



## Vernetztes Wissen – Daten, Menschen, Systeme

6. Konferenz der Zentralbibliothek, Forschungszentrum Jülich 5. - 7. November 2012

Bernhard Mittermaier (Hrsg.)

Proceedingsband

**WissKom**  
2012

 **JÜLICH**  
FORSCHUNGSZENTRUM





Forschungszentrum Jülich GmbH  
Zentralbibliothek

# **Vernetztes Wissen – Daten, Menschen, Systeme**

6. Konferenz der Zentralbibliothek  
Forschungszentrum Jülich

Bernhard Mittermaier (Hrsg.)

5. – 7. November 2012  
Proceedingsband

Schriften des Forschungszentrums Jülich  
Reihe Bibliothek / Library

Band / Volume 21

ISSN 1433-5557

ISBN 978-3-89336-821-1



Bibliografische Information der Deutschen Nationalbibliothek.  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der  
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten  
sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

Wir danken dem Hauptsponsor



Herausgeber  
und Vertrieb: Forschungszentrum Jülich GmbH  
Zentralbibliothek, Verlag  
D-52425 Jülich  
Telefon (0246 1) 61-5368 · Telefax (0246 1) 61-6103  
E-Mail: [zb-publikation@fz-juelich.de](mailto:zb-publikation@fz-juelich.de)  
Internet: <http://www.fz-juelich.de/zb>

Erstellung: Heike Lexis

Umschlaggestaltung: Grafische Medien, Forschungszentrum Jülich GmbH

Druck: Grafische Medien, Forschungszentrum Jülich GmbH

Grafik: Fotolia

Copyright: Forschungszentrum Jülich 2012

Schriften des Forschungszentrums Jülich  
Reihe Bibliothek / Library Band / Volume 21

ISSN 1433-5557  
ISBN 978-3-89336-821-1

Vollständig frei verfügbar im Internet auf dem Jülicher Open Access Server (JUWEL)  
unter <http://www.fz-juelich.de/zb/juwel>

Alle Rechte vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie oder  
in einem anderen Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder  
unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Konferenzkomitees .....	3
Vorwort .....	5
<b>Festvortrag</b>	
<b>Die Tugend des Vergessens in digitalen Zeiten .....</b>	<b>9</b>
Viktor Mayer-Schönberger	
<b>Vernetztes Wissen - Daten</b>	
<b>Zum Aufbau neuer Bibliotheksinfrastrukturen: Ein Bibliotheksverbund positioniert sich .....</b>	<b>17</b>
Silke Schomburg	
<b>Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften?</b>	
<b>Eine Analyse mit Fokus auf die Wirtschaftswissenschaften .....</b>	<b>29</b>
Sven Vlaeminck, Olaf Siegert	
<b>DOI® Registrierung sozial- und wirtschaftswissenschaftlicher Forschungsdaten mit da ra .....</b>	<b>45</b>
Brigitte Hausstein, Stefanie Grunow	
<b>re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositories. Ein Werkstattbericht .....</b>	<b>61</b>
Heinz Pampel, Hans-Jürgen Goebelbecker, Paul Vierkant	
<b>TEODOOR: Geodateninfrastruktur zur Verwaltung und Veröffentlichung von terrestrischen Beobachtungsdaten der HGF Infrastrukturmaßnahme TERENO aus verteilten Quellen .....</b>	<b>75</b>
Ralf Kunkel, Jürgen Sorg, Rainer Gasche, Jens Klump, Olaf Kolditz, Mark Frenzel, Frank Neidl	
<b>Auf kreativen Wegen von Daten zum Wissen am Beispiel medizinischer Forschungsdaten .....</b>	<b>93</b>
Lars Müller, Thomas Wetzel, Hans-Christoph Hobohm	
<b>Linked Data als Infrastruktur zur Integration von Forschungsdaten und Publikationen .....</b>	<b>107</b>
Dominique Ritze, Kai Eckert	
<b>Semantische Digitale Bibliotheken -</b>	
<b>Bibliotheken als Mehrwertproduzenten in vernetzten Wissensarchitekturen.....</b>	<b>119</b>
Stefanie Sorge	
<b>Vernetztes Wissen - Systeme</b>	
<b>Metadaten aus der Cloud: Technologien und Anwendungsfälle der Medienschließung mittels Software as a Service .....</b>	<b>135</b>
Michael Eble, Sebastian Kirch	
<b>Open-Access-Netzwerk: Semantische Anreicherung und Vernetzung von wissenschaftlichen OA-Inhalten .....</b>	<b>149</b>
Sammy David, Julia Iwanowa, Maxi Kindling, Michaela Voigt	
<b>Kooperation und Vernetzung als Kernaufgabe disziplinspezifischer Versorgung mit Volltexten: Das Fachrepository „OstDok“ .....</b>	<b>165</b>
Doris Škarić	
<b>Das Vernetzen von Menschen, Daten und Systemen –</b>	
<b>Die Forschungsbibliothek des Herder-Instituts in Marburg.....</b>	<b>179</b>
Jürgen Warmbrunn	
<b>Vernetzung von Datenbanken geisteswissenschaftlicher Forschungsprojekte .....</b>	<b>195</b>
Frank Förster, Michael Nolde, Bernhard Thalheim	

## Inhaltsverzeichnis

---

<b>Impact Messung – Mehrwertdienste für Repositorien .....</b>	<b>207</b>
Daniel Beucke	
<b>Spezifische Suchmaschinen mit der Open Source Suchmaschine YaCy am Beispiel der Johannes Gutenberg Universität Mainz .....</b>	<b>219</b>
Jörg Berkemeyer, Georg von Nessler	
<b>Vernetzung mit MyCoRe – Eine Repository-Software vernetzt Systeme, Daten und Menschen .....</b>	<b>225</b>
Wiebke Oeltjen	
<b>10 years collaborative catalogue enrichment - the dandelon.com community.....</b>	<b>235</b>
Manfred Hauer	
<b>Netzwerkarbeit in der Fraunhofer-Gesellschaft.....</b>	<b>247</b>
Andrea Siegberg	
<b>Vernetzung biographischer Online-Ressourcen – Catalogus Professorum Rostochiensium und Rostocker Matrikelportal.....</b>	<b>259</b>
Robert Stephan, Karsten Labahn, Matthias Glasow	
<b>Zur Entwicklung eines Linked-Open-Data-Dienstes für Bibliotheksdaten.....</b>	<b>271</b>
Felix Ostrowski, Adrian Pohl	
<b>Vernetztes Wissen - Menschen</b>	
<b>Das Netzwerk und die Institution .....</b>	<b>295</b>
Lambert Heller	
<b>Kollaborative Wissensgenerierung im virtuellen Raum – Entwicklungen und Erkenntnisse aus der internationalen Bildungsmedienforschung.....</b>	<b>307</b>
Sylvia Brink, Andreas L. Fuchs, Roderich Henrÿ, Robert Strötgen	
<b>Fachspezifische Anpassungen einer pandisziplinären Forschungsplattform am Beispiel der Pathologie .....</b>	<b>323</b>
Miroslaw Mainusch, Dominik Schmitz, Alberto Perez Bouza, Saskia Eva Elisabeth Westphal, Bernd Decker	
<b>Vom „embedded“ zum „liaison librarian“ – Was versprechen die neuen Konzepte? .....</b>	<b>337</b>
Simone Fühles-Ubach	
<b>24/7/365 X 360 – die neue Zauberformel?.....</b>	<b>351</b>
Anne-Katharina Weilenmann	
<b>Autorenliste.....</b>	<b>367</b>
<b>Sponsoren und Aussteller.....</b>	<b>379</b>

## Konferenzkomitees

### Programmkomitee

Dr. Ulrike Eich	Hochschulbibliothek, RWTH Aachen
Dr. Anne Lipp	Wissenschaftliche Literaturversorgungs- und Informationssysteme, Deutsche Forschungsgemeinschaft (DFG)
Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek
Heinz Pampel	Bibliothek und Informationsdienste LIS, Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum (GFZ)

### Organisationskomitee

Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek (Konferenzleitung)
Hilde Dobbstein	Forschungszentrum Jülich, Zentralbibliothek
Charlotte Esser	Forschungszentrum Jülich, Zentralbibliothek
Yasmin Fattah	Forschungszentrum Jülich, Zentralbibliothek
Dr. Christoph Holzke	Forschungszentrum Jülich, Zentralbibliothek

---

## **Vernetztes Wissen – Daten, Menschen, Systeme**

Der sprichwörtliche „Gelehrte im stillen Kämmerlein“ ist fast völlig verschwunden. Allenthalben gibt es interdisziplinäre Arbeitsgruppen, internationale Kooperationen und virtuelle Forschungsumgebungen. Grundlage solcher Arbeitsformen ist stets eine Vernetzung, die unter vielen verschiedenen Aspekten erfolgen kann. Allen Ansätzen ist gemeinsam, dass die Verknüpfung von Fragmenten etwas Neues hervorbringt, das in den isolierten Teilsystemen noch nicht vorhanden war.

Mit der WissKom2012 „Vernetztes Wissen – Daten, Menschen, Systeme“ greift die Zentralbibliothek des Forschungszentrums Jülich erneut Themen im Spannungsfeld von „Bibliothek - Information - Wissenschaft“ in einer Konferenz interdisziplinär auf und versucht, neue Handlungsfelder für Bibliotheken aufzuzeigen. Diese sechste Konferenz der Zentralbibliothek thematisiert den immer wichtiger werdenden Bereich der Forschungsdaten und den nachhaltigen Umgang mit ihnen. Sie zeigt auf, was Interdisziplinarität konkret bedeutet und wie bislang isolierte Systeme vernetzt werden können und so Mehrwert entsteht. Der Konferenzband enthält neben den Ausführungen der Referenten zudem die Beiträge der Poster Session sowie den Festvortrag von Prof. Viktor Mayer-Schönberger mit dem Titel „Delete: Die Tugend des Vergessens in digitalen Zeiten“.

Ich danke allen herzlich, die zum Gelingen der Konferenz beigetragen haben: Vortragende und Moderatoren, Aussteller, Mitglieder des Programmkomitees, Organisatoren und last but not least die Teilnehmerinnen und Teilnehmer. Ein besonderer Dank gilt dem Forschungszentrum Jülich und den Sponsoren, die diese Konferenz erst möglich gemacht haben.

Dr. Bernhard Mittermaier  
Forschungszentrum Jülich, Zentralbibliothek

---

# **Festvortrag**





## **Die Tugend des Vergessens in digitalen Zeiten**

Viktor Mayer-Schönberger

University of Oxford

Seit jeher vergessen wir Menschen das meiste, was wir täglich erleben, denken und fühlen. Das Merken und Erinnern hingegen fällt uns schwer. Wir können wenig dagegen tun – der Prozess des Vergessens ist einer, den wir bewusst nur in engen Grenzen kontrollieren können.

Deshalb haben wir Menschen seit Jahrtausenden versucht, Informationen die uns wichtig waren, dem Vergessen zu entziehen. Die Sprache gab uns die erste Möglichkeit, Dinge so festzuhalten, dass wir sie anderen einfach weitergeben konnten, über Distanzen und Generationen hinweg. So entstanden nicht zuletzt die großen Epen der Welt.

Mit Hilfe der Malerei versuchten unsere Vorfahren Erlebtes und Erdachtes zu fassen, damit es sich anders als gesprochenes Wort mit der Zeit nicht verändert. Aber die Malerei versagt, wenn es darum geht Informationen präzise über lange Zeit festzuhalten und zu erinnern. Das gelang mit Hilfe der Schrift – geboren einst aus der Notwendigkeit früher Buchhalter Dinge und Mengen festzuhalten.

Sprache, Malerei und Schrift half uns Erlebtes und Gedachtes zu externalisieren, und so die Grenzen unseres eigenen Gedächtnisses zu überwinden. Aber diese Werkzeuge des Erinnerns änderten nichts am Grundsätzlichen, dass für uns Menschen das Vergessen einfach ist, und das Erinnern aufwändig, schwierig, und kostspielig.

Daran änderte auch die Erfindung der Druckerpresse nichts, und der Phonograph, die Fotografie oder der Film. Das Erinnern auch mittels externer Erinnerungswerkzeuge blieb teuer, und war damit vor allem besonderen Ereignisse oder Gedanken vorbehalten.

Im digitalen Zeitalter hat sich dieses Verhältnis von Vergessen und Erinnern umgekehrt. Heute ist das Festhalten von Informationen in digitalen Erinnerungsspeichern, aber auch deren Auffindung so einfach und billig geworden, dass unsere digitalen Hilfsmittel heute in der Regel Informationen festhalten, speichern, und verfügbar halten. Das Vergessen hingegen bedarf eines bewussten Aktes, und ist zur Ausnahme geworden.

Diese Umkehrung von Vergessen und Erinnern im digitalen Zeitalter hat meines Erachtens eine Reihe von bedeutenden negativen Konsequenzen zur Folge, die – wenn sie unbeachtet bleiben – die Vorteile digitaler Speicher mehr als aufwiegen könnten.

Diese Nachteile lassen sich in zwei Kategorien einteilen, die ich mit den Schlagworten Macht und Zeit bezeichnen möchte.

(a) Macht: Datenschutzexperten warnen seit vielen Jahren, dass ein informationelles Machtgefälle, etwa ausgelöst dadurch, dass jemand über einen anderen etwas weiß, was der Betroffene selbst schon längst vergessen hat, unser Verhalten und unsere Entscheidungen prägen kann.

Aber es geht um mehr als relationale Machtgefälle zwischen zwei Transaktionspartnern. Erinnern wir uns nur daran, dass die Katholische Kirche viele Jahrhunderte lang ihre Macht nicht zuletzt darauf stützte, die gesellschaftlichen Institutionen des Erinnerns – von den Kopisten und Bibliotheken bis zu den Schulen – zu kontrollieren.

Hinzu kommt, dass schon der Glaube beobachtet zu werden ausreichen kann, um unser Verhalten zu beeinflussen. Bentham schlug einst ein Gefängnis vor, in dem die Gefangenen nicht wüssten ob und wann sie kontrolliert werden. Benthams Kalkül: Weil sie damit rechnen müssten, stets beobachtet zu sein, würden sie sich fortwährend wohlverhalten. Er nannte seine Erfindung das Panopticon.

Oscar Gandy und andere haben argumentiert, dass durch das Internet wir in einem globalen Panopticon leben, in dem wir stets davon auszugehen haben, dass die ganze Welt sieht, was wir online tun oder kommunizieren.

Durch das umfassende digitale Erinnern wird die Gefahr noch viel grösser: Wir müssen davon ausgehen, dass das, was wir tun, nicht nur im Heute wahrgenommen wird, sondern auch Jahre, ja Jahrzehnte in der Zukunft zugänglich bleibt – und dann aus dem zeitlichen Zusammenhang gerissen uns zum Nachteil wird. Das wiederum lässt uns, werden wir uns der Gefahr des digitalen Erinnerns bewusst, vorsichtig werden, uns nicht mehr in den Diskurs der Öffentlichkeit kritisch einmischen, weil wir fürchten, es würde uns in der Zukunft Nachteile bringen. Damit aber unterspülen wir ein zentrales Fundament unserer Demokratie.

(b) Zeit: Das menschliche Vergessen hilft uns auch mit alten und irrelevant gewordenen Informationen umzugehen: im Regelfall können wir uns dann an diese einfach nicht mehr erinnern. (Dabei lasse ich bewusst die Frage offen, ob der Vorgang des Vergessens Gemerktes tatsächlich tilgt oder nur den Verweis auf Gemerktes.) Bildlich gesprochen entsorgen wir so Erinnerungsmüll.

Weil diese Vergessensvorgänge aber ohne unsere bewusste Entscheidung ablaufen, haben wir Menschen auch keine bewussten kognitiven Fähigkeiten entwickelt zu vergessen. Das bedeutet, dass es uns schwer fällt mit Zeit als Dimension der Veränderung gerade in der Vergangenheit umzugehen. Wir kennen das Phänomen aus dem Alltag: werden wir durch einen externen Stimulus (etwa ein altes Bild, einen alten Brief) an etwas erinnert, das wir schon fast vergessen wähnten, empfinden wir manchmal die Erinnerung wieder sehr frisch – und mitunter stärker als später Erlebtes, aber jetzt nicht Erinnertes. Das bereitet uns Schwierigkeiten unterschiedliche Erfahrungen in einen rationalen zeitlichen Zusammenhang zu stellen, wenn wir etwa schon lange vergessen Geglaubtes plötzlich intensiv erinnern.

Wissenschaftler haben in diesem Zusammenhang einige Menschen untersucht, die Schwierigkeiten haben zu vergessen. Sie erinnern sich auch Jahre später an jedes Detail eines jeden Tages. Diese Menschen sehen ihr umfassendes Gedächtnis nicht als Wohltat, sondern empfinden es als Last. Sie sagen, dass die umfassende Erinnerung es schwer macht in der Gegenwart zu entscheiden – weil ihnen alle Fehlentscheidungen der Vergangenheit stets präsent sind. Mehr noch: Indem wir vergessen vergeben wir. Ein Vergeben ohne Vergessen hingegen fällt uns viel schwerer.

Durch ein umfassendes digitales Erinnern könnten wir deshalb in Schwierigkeiten geraten im Hier und Jetzt zu entscheiden, aber auch uns selbst und anderen Vergangenes zu vergeben. Wir würden uns und andere die Chance nehmen uns zu verändern.

Aber mit dieser Art der digitalen Auslagerung von Erinnerung ist noch eine weitere Gefahr verbunden: Sie bemächtigt jene vorwiegend kommerzielle Unternehmen, die es sich zur Aufgabe gemacht haben unsere digitalen Informationen aufzuzeichnen und vorrätig zu halten – wie Google, Facebook, YouTube und Flickr. Diese haben nun nicht nur großes Wissen über uns, sondern auch die Macht unsere digitale Vergangenheit zu verändern.

Gleichzeitig werden die gesellschaftlichen Institutionen des Erinnerns – die Bibliotheken, Archive, Museen und Gedenkstätten – entmachtet. Die Politik streicht ihnen die Budgets mit Hinweis auf die „kostenlose“ digitale Speicherung durch kommerzielle Anbieter im Netz. Und aus dem immer grösser werdenden Meer an informationeller Beliebigkeit ragen diese Institutionen nicht mehr als Leuchttürme zur informationellen Orientierung heraus; sie verlieren vermeintlich an Bedeutung.

Das aber wäre ein schlimmes Ergebnis: es würde nicht nur den Institutionen sondern auch gerade jenen Informationen, die wir als Gesellschaft beschlossen haben festzuhalten, die Sichtbarkeit rauben, weil sie kaum mehr wahrgenommen werden.

Das sind nur einige der Gefahren, die uns drohen, wenn wir uns reflexionslos dem digitalen Erinnern ergeben.

Dem gegenüber stelle ich die Forderung nach einem erleichterten Vergessen gerade im digitalen Zeitalter: also der Verfügbarkeit von technischen, institutionellen, rechtlichen, und organisatorischen Maßnahmen, die das unreflektierte digitale Erinnern erschweren und das Vergessen – das Verlieren von Recall digitaler Daten – erleichtert. Mir geht es nicht darum, das Erinnern zu verunmöglichen, sondern nur jenes Verhältnis zwischen Erinnern und Vergessen wieder herzustellen, mit dem wir Menschen vertraut sind.

Eine pragmatische Möglichkeit dazu ist ein Verfallsdatum für Informationen: dieses wird wie andere Metadaten auch mit der Information selbst verknüpft und unsere technischen Hilfsmittel sind so eingestellt, dass „abgelaufene“ Information gelöscht oder jedenfalls nicht einfach wieder aufrufbar wird.

Das Verfallsdatum würde schon zum Zeitpunkt des Speicherns (bzw Uploads) eingegeben, und könnte vom Betroffenen jederzeit später auch geändert werden.

Mir geht es bei diesem Vorschlag nicht um technische Perfektion – diese halte ich für nicht erreichbar, sondern um einen pragmatischen Ansatz, der einen Gutteil der Probleme digitalen Erinnerns lösen könnte, wenn nur die großen fünf oder zehn Internet-Dienste mitmachen.

Vor allem aber geht es mir darum, dass wir in unserer Gesellschaft die Debatte darüber intensivieren, was und von wem wie lange digital „erinnert“ werden kann und soll. Denn damit schon beginnen wir zu begreifen, dass es sich hier um eine gesellschaftliche Frage handelt, die wir nicht den Entscheidungen kommerzieller Unternehmen alleine überlassen können. Lassen Sie uns gerade in der digitalen Zukunft erinnern zu vergessen!

---

# **Vernetztes Wissen – Daten**





## **Zum Aufbau neuer Bibliotheksinfrastrukturen: Ein Bibliotheksverbund positioniert sich**

Silke Schomburg

Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz)

Die Arbeitsweise von Wissenschaftlern hat sich in den vergangenen Jahren drastisch verändert und wandelt sich noch. Wissenschaftler sehen sich heute vielen neuen Werkzeugen und Verfahrensweisen gegenüber, die Forschung und Lehre effektiver werden lassen und bisher nicht zu erreichende Erkenntnisse ermöglichen. Die neuen Methoden lösen etablierte wissenschaftliche Verfahren ab oder ergänzen sie<sup>1</sup>. Vor allem die wissenschaftliche Kommunikation und die Informationsversorgung sind einem Wandel unterlegen, der insbesondere die Rolle und Aufgabe der Bibliotheken sowie ihrer Dienstleister betrifft. Bibliotheken und Bibliotheksverbünde reagieren auf die neuen Anforderungen und gestalten den Wandel offensiv mit. Die Rolle und die Aufgaben von Informationseinrichtungen sind dabei untrennbar an die Art und Weise wissenschaftlicher Publikation und Kommunikation geknüpft; die Effektivität wissenschaftlichen Arbeitens hängt unmittelbar von intakten und bezahlbaren Zugängen zu forschungsrelevanten Informationen ab. Spezialbibliotheken können spartenspezifische Lösungen für ihren Kompetenzbereich und ihre Wissenschaftler aufbauen, die Zusammenführung und Prüfung, ob daraus neue zentrale Dienste für die Wissenschaftlichen Bibliotheken aufgebaut werden können, obliegt den Bibliotheksverbünden.

Das Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz) ist als Einrichtung des Landes<sup>2</sup> eine von insgesamt sechs Verbundzentralen und zentraler Dienstleister für die Hochschulen in Nordrhein-Westfalen und den größten Teil von Rheinland-Pfalz. Die Produkte und Dienste des hbz kommen Bibliotheken und Wissenschaftlern zugute.

Verschiedenste Fragestellungen aus den Bereichen Information und Kommunikation werden durch das hbz bedient.

---

<sup>1</sup> Jim Gray (Microsoft Research) hat in diesem Sinn den Begriff des 4. Forschungsparadigmas geprägt. Die daten-intensive Wissenschaft bringt Daten, Methoden und Disziplinen zusammen, verlangt aber auch nach entsprechenden Technologien und Offenheit gegenüber Neuem (vgl. The Fourth Paradigm: Data-Intensive Scientific Discovery, hrsg.von Tony Hey, Stewart Tansley, Kristin Tolle, Publisher: Microsoft Research 2009)

<sup>2</sup> nach § 14 des Landesorganisationsgesetzes vom 10. Juli 1962 (SGV. NRW. 2005), Link zur Satzung: <http://www.hbz-nrw.de/ueberuns/satzung/>

Es unterstützt die Wissenschaftler unmittelbar bei Forschung und Lehre: Suchen sie nach Veröffentlichungen und neuesten Zeitschriftenartikeln, wollen sie eine Fernleihbestellung aufgeben oder eine neue elektronische Zeitschrift gründen oder möglicherweise ein Digitalisierungsprojekt an den Start bringen? Vermittelt durch die Bibliotheken der Region, erbringt das hbz zuverlässige Unterstützung und innovative Lösungen für all diese Aufgaben, ohne dass es als Dienstleister in den Angeboten prominent auftritt. Die Dienstleistungen der dem hbz angeschlossenen Verbundbibliotheken (und damit auch die des hbz) orientieren sich an den Interessen ihrer Endkunden, der Wissenschaftler und Studierenden. Der vorliegende Artikel beschäftigt sich mit den aktuellen Entwicklungen und neuen Aufgaben in der Informationsversorgung aus dem Blickwinkel des hbz.

## Rückschau

Das hbz hat sich 2009 intensiv mit der Planung von zukünftigen Diensten beschäftigt und Entwicklungen und Tendenzen im Bereich der wissenschaftlichen Information und Kommunikation herausgearbeitet. Das in diesem Zusammenhang erarbeitete "*Perspektivpapier*" stellt für das hbz den Einstieg in die Diskussion über die Hauptfelder unserer zukünftigen Produkte und Dienstleistungen dar. Im Kontext mit der Konferenz "*Digitale Wissenschaft*"<sup>3</sup> im September 2010 bot sich erneut die Gelegenheit für eine Positionsbestimmung des Bibliotheksverbundes zusammen mit Bibliothekaren und Wissenschaftlern. Damals wurde konstatiert, dass man sich im Bereich der schriftbasierten wissenschaftlichen Kommunikation bewegt und sich hauptsächlich mit institutionell gebilligten wissenschaftlichen Schriften befasst, jedoch noch nicht ausreichend z. B. mit Blog- und Wikibeiträgen oder Forschungsdaten, da all diese Publikationen bzw. Informationen noch keine ihnen zugestandene Zitierfähigkeit mit sich bringen. Im Rückblick zeigt sich hier bereits, die rasche Veränderung der Informationslandschaft. Wikis, Blogs und Forschungsdaten<sup>4</sup> sind zunehmend unverzichtbarer Bestandteil der wissenschaftlichen Arbeit.

Ebenfalls ab September 2010 startete die Evaluierung der Bibliotheksverbünde durch den Wissenschaftsrat und die Deutsche Forschungsgemeinschaft (DFG).

---

<sup>3</sup> Tagung "Digitale Wissenschaft, Köln, 20. und 21. September 2010, Link zur Publikation unter: [http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung\\_Digitale\\_Wissenschaft.pdf](http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung_Digitale_Wissenschaft.pdf)

<sup>4</sup> z.B. Textkorpora in der virtuellen Forschungsumgebung TextGrid: <http://www.textgrid.de/>

Die Verbünde hatten sich mit einem umfangreichen Fragenkatalog auseinander zu setzen, der nicht den üblichen Fragen der Wissenschaftsevaluierungen entsprach. Die Beschäftigung mit dem Fragenraster führte an vielen Stellen erneut zu Richtungsentscheidungen innerhalb der Bibliotheksverbünde und markierte den Beginn einer infolge des engen Zeitrahmens sehr intensiven und grundsätzlichen Zukunftsdiskussion aus verschiedenen Perspektiven; erst auf eigene Veranlassung, dann im Kontext mit der Wissenschaft und schließlich im Umfeld der Evaluierungen. Der durch die Evaluierung gegebene Rahmen hat den Diskurs über eine perspektivische Positionierung einerseits befördert, andererseits auch viel Zeit gekostet. Da die Umsetzung der Evaluierung der Bibliotheksverbünde durch den Wissenschaftsrat und die Deutsche Forschungsgemeinschaft (DFG) noch andauert, führte sie in den vergangenen beiden Jahren immer wieder zum Hinterfragen der Produktpalette und Neupositionieren der Angebote des hbz. Der stattfindende Wandel in der Informationsversorgung drückte sich dementsprechend auch direkt im hbz-Produktportfolio aus: Bisherige Dienste wurden ausgebaut, umorganisiert oder abgewickelt, da sich die Nachfrage in den Bibliotheken geändert hatte. Funktionalitäten wurden angepasst und erweitert und neue Fragestellungen (z.B. Open Linked Data) forciert angegangen.

Was bei den Hochschulbibliotheken in Nordrhein-Westfalen und anderswo nach dem Einsetzen der Hochschulfreiheit<sup>5</sup> konstatierbar ist – nämlich die Förderung der Konkurrenz zwischen den einzelnen Standorten und die fachliche Profilierung der Einrichtungen – ist bei den bibliothekarischen Verbünden in ihrer derzeitigen Aufstellung scheinbar nur schwer umzusetzen, da dem nicht zuletzt die Wünsche der Kunden nach regionalen Angeboten, regionalen Ansprechpartnern und maßgeschneiderten Lösungen im Wege stehen.

---

<sup>5</sup> Link unter:  
[https://recht.nrw.de/lmi/owa/br\\_bes\\_text?anw\\_nr=2&gld\\_nr=2&ugl\\_nr=221&bes\\_id=9796&aufgehoben=N&menu=1&sg=](https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=221&bes_id=9796&aufgehoben=N&menu=1&sg=)

Gerade dieses – die fachliche Differenzierung und der Abbau von Mehrfacharbeiten – war jedoch eine der zentralen Forderungen in den Empfehlungen von Wissenschaftsrat<sup>6</sup> und DFG<sup>7</sup>, die Anfang 2011 veröffentlicht wurden. In den kommenden Monaten haben sich die Verbünde und die regionalen Gremien<sup>8</sup> mit den Empfehlungen intensiv beschäftigt. Unter anderem spielt auch die politische Verankerung der Evaluierungsergebnisse eine zentrale Rolle für die zukünftigen Aktivitäten der Verbünde. In diese Diskussion "platzte" die Anfang des Jahres von der DFG veröffentlichte Ausschreibung<sup>9</sup> zur Neuausrichtung überregionaler Informations-services hinein und führte zu einer Konzentration der Verbünde auf die Beteiligung an der Ausschreibung.

### DFG-Ausschreibung

*Die DFG-Ausschreibung steuert zweifellos mit den von den DFG-Gremien festgelegten vier Schwerpunkten zukünftige Betätigungsfelder der Verbünde, in denen aus Sicht der DFG "die Entwicklung neuer Strukturen und Services am vordringlichsten erscheint"<sup>10</sup>. Dazu gehören:*

- Bibliotheksdateninfrastruktur und Lokale Systeme,
- Langzeitverfügbarkeit,
- Electronic Resource Management – das Management digitaler Publikationen,
- Forschungsnahe Informationsinfrastruktur.

---

<sup>6</sup> Wissenschaftsrat: „Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland“, veröffentlicht im Januar 2011, Link unter:

<http://www.wissenschaftsrat.de/download/archiv/10463-11.pdf>

<sup>7</sup> „Positionspapier zur Weiterentwicklung der Bibliotheksverbünde als Teil einer überregionalen Informationsinfrastruktur“, erarbeitet vom DFG-Ausschuss für wissenschaftliche Bibliotheken und Informationssysteme, veröffentlicht im Januar 2011, Link unter:

[http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier\\_bibliotheksverbuende.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_bibliotheksverbuende.pdf)

„Gemeinsame Erklärung der Deutschen Forschungsgemeinschaft und des Wissenschaftsrats zur Zukunft der Bibliotheksverbünde als Teil einer überregionalen Informationsinfrastruktur in Deutschland“, veröffentlicht im Februar 2011, Link unter:

[http://www.dfg.de/download/pdf/foerderung/programme/lis/gemeinsame\\_erklaerung\\_dfg\\_wr\\_bibliotheksverbuen\\_de.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/gemeinsame_erklaerung_dfg_wr_bibliotheksverbuen_de.pdf)

<sup>8</sup> In der Zwischenzeit positionierten sich Mitte 2011 die nordrhein-westfälischen Universitätsbibliotheken mit Überlegungen zu zentralen Diensten, die u.a. im Umfeld der Bibliotheksverbünde zu sehen sind und auch überregional im Rahmen der Sitzungen der Sektion 4. des Deutschen Bibliotheksverbundes (DBV) diskutiert wurden. Link unter:

[http://www.bibliotheksverband.de/fileadmin/user\\_upload/Sektionen/sektion4/Tagungen/2010\\_Zentrale\\_Dienste\\_f%C3%BCr\\_Hochschulbibliotheken.pdf](http://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Tagungen/2010_Zentrale_Dienste_f%C3%BCr_Hochschulbibliotheken.pdf)

[http://www.bibliotheksverband.de/fileadmin/user\\_upload/Sektionen/sektion4/Tagungen/2010\\_Zentrale\\_Dienste\\_f%C3%BCr\\_Hochschulbibliotheken.pdf](http://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Tagungen/2010_Zentrale_Dienste_f%C3%BCr_Hochschulbibliotheken.pdf)

<sup>9</sup> Ausschreibungstext veröffentlicht am 12. Januar 2012:

[http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung\\_ueberregionale\\_informationsservices\\_121015.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_ueberregionale_informationsservices_121015.pdf)

<sup>10</sup> Stand 26.9.2012:

[http://www.dfg.de/foerderung/info\\_wissenschaft/info\\_wissenschaft\\_12\\_03/index.html](http://www.dfg.de/foerderung/info_wissenschaft/info_wissenschaft_12_03/index.html)

Im hbz folgte eine intensive Diskussion zu dieser Ausschreibung, in der im Kontext zur Ausschreibung der Neuausrichtung überregionaler Informationsinfrastrukturen erneut die Fragen

- wo wollen wir uns zukünftig engagieren,
- wo ist Neues zu entwickeln bzw. müssen Partnerschaften eingegangen werden,
- wo sehen wir Bedarf für neue Angebote,
- wo wollen wir vorangehen,

eine zentrale Rolle spielten. Dabei sind vier Prämissen zu berücksichtigen, die bei den Planungen und bei der Beantwortung der Fragen beachtet werden müssen, denn Open Source, Open Content, Open Data und Open Access sind für das Land NRW und für das hbz von strategischer Bedeutung<sup>11</sup>. Sie stellen nach Ansicht des Landes und des hbz die wesentlichen Bausteine für eine zukunftsfähige und bedarfsgerechte Informationsversorgung in einer Wissensgesellschaft im Umbruch dar, in der Daten eine immer wichtigere Rolle spielen, gleichzeitig ihren Wert aber nur behalten können, wenn es Mittel und Wege gibt, diese Flut von Informationen zu sichten, zu bewerten, zu verwalten und zu verarbeiten.

Das hbz tritt deshalb in der Region nicht nur als Dienstleister der Bibliotheken, sondern verstärkt auch als Kompetenz- und Beratungszentrum an. Die anstehenden Veränderungen beim Auf- und Umbau der wissenschaftlichen Informationsinfrastruktur machen eine solche Unterstützung notwendig. Der schon stattfindende Wandel der Bibliotheken von der Fokussierung auf den lokalen Bestand hinaus hin zur konsequenten Nutzerorientierung mit zukunftsweisenden Informationsarchitekturen und Zugangssystemen zu digitalem Content mit neuen Formen der Erschließung wird vom hbz aktiv begleitet.

Mit Hilfe eines Versuchs einer Gesamtschau, einer Auflistung der unterschiedlichsten Aufgaben an Bibliotheken, die durch die Verbünde erbracht werden bzw. aus unserer Sicht zukünftig erbracht werden können, kann eine Aufgliederung in acht Bereiche vorgenommen werden, die die Arbeit der Bibliotheken in unterschiedlicher Ausprägung bestimmen:

---

<sup>11</sup> Zur Weiterentwicklung von Systemen: Das hbz wählt OS-Systeme aus unterschiedlichen Bereichen, wenn keine anderen attraktiven und bezahlbaren Angebote vorhanden sind oder im Fall von Nischenprodukten, wie z.B. die Deutsche Bibliotheksstatistik (DBS); nach reiflicher Überlegung wird im hbz eine Neuprogrammierung ab 2013/14 angegangen.

- *Ausleihe und Benutzung*  
(Identity-Management-Systeme, Virtuelle Auskunft, Fernleihsteuerung, Dokumentlieferdienste),
- *Erwerbung*  
(Konsortiale Erwerbung, Profildienste, ERM, Patron Driven Aquisition, Neuerwerbungsdienste, Schnittstellen zu/Integration von Hochschulinfrastruktur),
- *Recherche*  
(Index/Lokaler Suchraum, Informationsportale),
- *Erschließung/Katalogisierung*  
(Katalogisierungsdatenbank, Sacherschließung, Kataloganreicherung, Linkresolver, Retro-Katalogisierung, Fremddaten, Normdaten, Datenversorgung der lokalen Systeme, Semantic-/Social-Web-Anwendungen),
- *Content Management*  
(Infrastruktur und Workflow, Hosting von Datenbanken/Publikationen, Digitalisierung, LZA, Forschungsdaten, Open-Access-Dienste),
- *Publishing*  
(Workflow für Publishing, Hochschulbibliografie, Bibliometrische Dienste, Open Access, Forschungsdaten)
- *E-Learning*  
(E-Semesterapparate, E-Learning-Plattformen, Informationskompetenz),
- *Fort- und Weiterbildung.*

Die aufgeführten Arbeitsbereiche sind für die Dienstleistungen der Verbünde unterschiedlich bedeutsam und bieten an vielen Stellen Handlungsoptionen, die vom hbz strategisch gewichtet wurden. Dazu zählen z. B. Schnittstellen<sup>12</sup> und Hosting-Dienste<sup>13</sup>, die schon mit Hilfe der Fachabteilungen und des hauseigenen Fachrechenzentrums angeboten und ausgebaut werden oder der schon begonnene Aufbau von Basisinfrastrukturen im Bereich Linked Open Data<sup>14</sup>.

Die (nicht nur von uns postulierten) Vorteile von zentralen Hosting-Diensten müssen nicht in jedem Fall auf eine kommerzielle Cloud-Lösung hinauslaufen, wie es derzeit die beiden großen Systemanbieter (Alma<sup>15</sup> und WorldShare<sup>16</sup>) planen. Ein zuverlässiger Verbleib und eine Nachnutzung der bibliothekarischen Daten im öffentlich-rechtlichen Raum sowie eine Verfügbarkeit von wissenschaftlich relevanten Informationen auch diesseits der hochpreisigen Schranken der Verlage und Datenbankanbieter sind Hauptaufgaben, die das hbz erbringen möchte. Die nationale Datenhoheit im öffentlichen Raum hat dabei eine besondere Bedeutung

---

<sup>12</sup> Mit dem baukastenartigen Aufbau der Produkte wird z.B. die Integration mit Angeboten Dritter einfacher. (Beispiel: Ausbau von Schnittstellen zwischen der Verbunddatenbank und Drittanwendungen mit Born Digital Dokumenten), siehe: <http://www.pressebox.de/attachment/510746/Anwenderbericht-Paderborn.pdf>

<sup>13</sup> <http://www.hbz-nrw.de/angebote/hosting/>

<sup>14</sup> [http://www.hbz-nrw.de/projekte/linked\\_open\\_data/](http://www.hbz-nrw.de/projekte/linked_open_data/)

<sup>15</sup> <http://www.exlibrisgroup.com/category/AlmaOverview>

<sup>16</sup> <http://www.oclc.org/de/de/webscale/default.htm>

unter der Prämisse, dass wir heute noch nicht alle Anwendungen und Dienste kennen können, die auf Basis dieser mit überwiegend öffentlichen Mitteln entstandenen Daten möglich sind und von der Wissenschaft benötigt werden. Deshalb sind die Konzepte von Linked Open Data und Open Content wesentlich für die moderne Informationsversorgung.

Zu diesen gewählten Aufgaben passen von den unterschiedlichen Feldern der DFG-Ausschreibung die folgenden gut: Bibliotheksdateninfrastruktur und Lokale Systeme, Langzeitverfügbarkeit und Electronic Resource Management – das Management digitaler Publikationen. Das hbz hat sich entsprechend in den drei Feldern mit der Beteiligung an einem Antrag positioniert und plant zusätzlich zusammen mit zwei weiteren Verbünden, die beiden wichtigsten Open-Source-Lokalsysteme koha<sup>17</sup> und Kual<sup>18</sup> zu evaluieren, um den Bibliotheken bei Bedarf ein entsprechendes Angebot machen zu können.

### **Bibliotheksdateninfrastruktur**

Die Forderung nach einer zentralen, offenen und herstellerunabhängigen Erschließungs- und Nachweisplattform für bibliographische Metadaten wird basierend auf erprobten Komponenten mit dem bei der DFG beantragten Projekt libOS (Library Operating System) bedient. libOS soll die Basis einer zukunftsfähigen und flexiblen Informationsinfrastruktur in Deutschland bilden sowie die parallele Katalogisierung aufheben und die in den Verbünden verteilten Datenbestände zusammenführen. Das Projekt gewährleistet die öffentliche Kontrolle der wissenschaftlichen Informationsversorgung, fördert den Wettbewerb durch das Aufbrechen bestehender Monopole und trägt so entscheidend dazu bei, die Innovationsfähigkeit und Exzellenz der Forschung sowie Lehre in Deutschland zu sichern und auszubauen.

Mit der Bereitstellung einer Entwicklerplattform, die auf einer offenen, einfach nutzbaren Service-API basiert, sowie die Publikation aller Daten als Linked Open Data schaffen die beteiligten Einrichtungen die Grundlage für Erweiterungen und Integration in Dienstleistungen, die wahlweise durch Bibliotheken, Wissenschaftler und andere Akteure vorgenommen werden können. Auf diese Weise soll libOS langfristig das Rückgrat eines weit vernetzten Informationsökosystems bilden.

---

<sup>17</sup> <http://www.koha.org/>

<sup>18</sup> <http://www.kuali.org/>



### **Langzeitverfügbarkeit**

Bestrebungen zur Umsetzung einer umfassenden Digitalen Langzeitarchivierungslösung in Deutschland sind nicht neu. Es gibt eine ganze Reihe von Systemen, die partielle Lösungen anbieten und ihrer Struktur nach entweder regional oder in ihren Disziplinen so fest verankert sind, dass es kaum anzunehmen – und auch nicht wünschenswert – ist, sie durch eine zusätzliche nationale Lösung zu ersetzen. Andererseits hat die Koexistenz mehrerer Lösungen nationaler Reichweite offensichtliche Nachteile, die um so gravierender sind, je näher ein Nutzer an den inhaltlichen Anwendungen und je weiter er von der institutionell-infrastrukturellen Sicht entfernt ist.

Mit seinem Vorhaben im Rahmen der DFG-Ausschreibung will das hzb gemeinsam mit seinen Partnern die Grundlagen für eine belastbare und zukunftsfähige Langzeitverfügbarkeitslösung aufbauen, die die unterschiedlichen Bedürfnisse der Informationseinrichtungen im wissenschaftlichen Kontext erfüllt. Die geplante Infrastruktur stellt ein überregionales Angebot für die Archivierung digitaler Materialien dar (z. B. born digitals, retrokonvertierte Materialien unterschiedlicher Provenienz) dar. Durch die Anbindung und Bereitstellung bereits existierender bzw. im Aufbau befindlicher Ansätze zur Langzeitverfügbarkeit kann eine Vielzahl spezifischer Lösungen angeboten werden, ohne dass diese neu entwickelt werden müssen.

Das Vorhaben unternimmt im Gegensatz zu anderen Projekten, nicht den Versuch eine weitere Komplettlösung für den Bereich der Langzeitverfügbarkeit zu erfinden. Vielmehr soll eine einheitliche Infrastruktur zur Bedienung und Auswahl bereits existierender Systeme geschaffen werden. Der Fokus des Projektes liegt damit auf der Integration existierender komplexer Systeme und nicht auf der Neuentwicklung ebensolcher. Die Infrastruktur ist damit leichtgewichtig und im Unterschied zu einer monolithischen nationalen Lösung leicht umzusetzen, ohne dass auf Merkmale einer solchen Lösung verzichtet werden muss.

### **Electronic Resource Management – das Management digitaler Publikationen**

Das Projekt baut ein bundesweit verfügbares Electronic Resource Managementsystem (ERMS) auf, das Bibliotheken und Konsortialstellen auf der

Basis einer zentralen Knowledge Base eine einheitliche Nutzung von Daten zur Lizenzverwaltung elektronischer Ressourcen auf lokaler, regionaler und nationaler Ebene ermöglicht. Das modular aufgebaute System verfügt über ein Zugangsmanagement sowie über geeignete Schnittstellen, um Statistikserver zu integrieren und eine komfortable Auswertung statistischer Daten zur Nutzung lizenzierter elektronischer Medien zu ermöglichen. Vor allem Bibliotheken beim bedarfsgerechten Einsatz von Budgets für eine kontinuierlich wachsende Anzahl elektronischer Ressourcen unterstützt werden. Offene Schnittstellen gekoppelt mit dem Zugangsmanagement machen das ERMS in andere Systeme integrierbar und verbinden diese Dienste mit denen der Bibliotheksdateninfrastruktur.

### **Weitere neue Infrastrukturen**

Wie bereits erwähnt, hat sich das hzb schon vor der Evaluierung und DFG-Ausschreibung mit der Weiterentwicklung seiner Dienstleistungen für eine kundengerechte Informationsversorgung beschäftigt. Wir haben auf "konventionelle" Art die Recherche- und Nachweissysteme weiter optimiert. Egal, ob es sich dabei um den Einsatz von Suchmaschinentechnologie, die Verbesserung der elektronischen Fernleihe oder um Digitalisierungsprojekte handelt, immer war es auf die eine oder andere Art Kataloganreicherung und -optimierung und damit war es schlichtweg nur eine Verbesserung der eigenen Angebote. Für eine verbesserte Recherche war das zugegebenermaßen optimal, aber nicht förderlich im Sinne der Nachnutzung unserer Daten durch Dritte, seien es Wissenschaftler oder z. B. kommerzielle Suchmaschinen. Die dauerhafte und gute Sichtbarkeit und Präsenz der Daten und Institutionen im Netz sind eines der wichtigsten Ziele der Zukunft. Lange wurde der Aspekt der möglichst einfachen Nachnutzung ausgeblendet, indem alle Angebote in Katalogen verschlossen blieben. Um dieses Ziel zu erreichen, gingen wir gleichzeitig in drei Schritten vor: Die Freigabe der bibliothekarischen Metadaten unter CC0<sup>19</sup>, der Aufbau eines Exports in RDF und das Engagement im Projekt Culturegraph zusammen mit der Deutschen Nationalbibliothek.

Eine Freigabe der Katalogdaten unter CC0 und der damit verbundene Verzicht auf jegliche Einschränkung der Verwendung der Daten maximiert die Möglichkeiten der Nachnutzung für experimentelle und innovative Anwendungen auf Basis der

---

<sup>19</sup> <http://creativecommons.org/publicdomain/zero/1.0/>

Katalogdaten<sup>20</sup>. Das hbz ist mit diesem rechtlichen Ansatz auf Basis von CC0 Teil einer wachsenden Gemeinschaft von Einrichtungen, die ihre Daten offen stellen, zu der unter anderem Europeana, Wikipedia, die British Library und weitere Nationalbibliotheken zählen<sup>21</sup>.

Auf technischer Ebene wird eine Nachnutzung der Katalogdaten durch drei sich ergänzende Wege ermöglicht:

- via http-URI Dereferenzierung inklusive Content Negotiation für diverse Serialisierungen, darunter RDFa (in HTML eingebettetes RDF),
- 2. als RDF-Vollabzug,
- 3. über die Bereitstellung einer auf einem Triplestore basierenden SPARQL-Schnittstelle.

Diese drei Publikations-Technologien ermöglichen die maschinenlesbare Konsumtion der Katalogdaten auf vielfältige Weise und unterstützt damit ihre Integration in das Web und das Semantic Web. Dies gestattet etwa eine Indexierung der hbz-Daten durch Google<sup>22</sup>, eine Integration unserer ortsbezogenen Daten in externe Dienste<sup>23</sup>, oder die Anzeige von Inhaltsverzeichnissen in der ISBN-Suche der Wikipedia<sup>24</sup>.

Im Projekt Culturegraph baut das hbz mit der DNB einen Resolving- und Lookup-Dienst für bibliothekarische Identifikatoren auf<sup>25</sup>. Die Nutzung solcher gemeinsamer, globaler Identifikatoren ermöglicht eine stärkere Sichtbarkeit von Bibliotheken und ihren Beständen im World Wide Web. Daneben engagiert sich das hbz im Projekt Culturegraph beim Aufbau einer Open-Source-Infrastruktur (etwa für Code-Reviews und kontinuierlichen Integration der verteilten Entwicklungsarbeiten) zur Etablierung einer offenen technischen Community für Gedächtnisinstitutionen<sup>26</sup>.

Das hbz geht davon aus, dass die bibliothekarische Infrastruktur ein wichtiger Teil einer zunehmend spartenübergreifenden Infrastruktur ist. Der Aufbau neuer innovativer Wissenssysteme, seien es virtuelle Forschungsumgebungen, Forschungsnetze oder Portale zur Präsentation des digitalen Kulturerbes wie die Deutsche Digitale Bibliothek (DDB) und Europeana zeigt, dass der Eingangs des

---

<sup>20</sup> <http://creativecommons.org/weblog/entry/33768>

<sup>21</sup> [http://wiki.creativecommons.org/CC0\\_use\\_for\\_data](http://wiki.creativecommons.org/CC0_use_for_data)

<sup>22</sup> <https://wiki1.hbz-nrw.de/display/SEM/2012/06/08/Google+indexiert+Teile+des+LOD+Katalogs>

<sup>23</sup> <http://lodum.de/post/28619267432/linking-bibliographic-resources>

<sup>24</sup> <http://de.wikipedia.org/wiki/Spezial:ISBN-Suche/382741685X>

<sup>25</sup> <https://wiki1.hbz-nrw.de/display/SEM/Resolving+und+Lookup-Dienst+fuer+bibliothekarische+Identifizier+in+culturegraph.org>

<sup>26</sup> <https://github.com/culturegraph>

Textes erwähnte Wandel auch mit einer starken Vernetzung von Informationsressourcen und -technologien einhergeht. Zur Verstetigung der bibliothekarischen Infrastruktur in diesem Kontext ist eine enge Zusammenarbeit der Verbundfachrechenzentren zu organisieren wie es schon spartenübergreifend in NRW mit einem OAIS kompatiblen landesweiten Langzeitarchiv im Projekt Digitales Archiv Nordrhein-Westfalen (DANRW) geschieht.

In den vom Programmkomitee ausgewählten Beiträgen lässt sich aus Sicht der Vortragenden Einrichtungen einiges (neue Felder, schon begonnene Themenstellungen, noch offene Fragen) ablesen: Was ist das? Und was fehlt noch? Woher bekommt man die Lobby, um dies umzusetzen? Woher die dafür notwendige Finanzierung, etwaige Kooperationspartner und das entsprechend gerüstete und motivierte Personal? Was könnten überregionale, nationale und internationale Projekte bewegen? Wohin entwickeln sich die kommerziellen Dienstleister? Stellen Sie uns maßgeschneiderte Lösungen für die Zukunft zur Verfügung oder sind sie mehr auf kurzfristigen Shareholder-Value ausgerichtet sowie auf Know-how-Abbau in den öffentlich geförderten Einrichtungen? Welche Dinge kann man abgeben und welches Know-how muss erhalten bleiben, um nicht in zu große Abhängigkeit von kommerziellen Herstellern zu geraten?

Ich freue mich darauf, auf der Konferenz einen Teil der Fragen beantwortet zu bekommen und für andere Denkansätze und neue Konzepte mitnehmen zu können.

---

## **Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften**

Sven Vlaeminck, Olaf Siegert

**ZBW - Leibniz Informationszentrum Wirtschaft**

### **Zusammenfassung**

In den Wirtschafts- und Sozialwissenschaften nehmen empirische Untersuchungen an Bedeutung zu. Dies schlägt sich auch in einer steigenden Anzahl von entsprechenden Beiträgen in Fachzeitschriften nieder, bei denen Autoren<sup>1</sup> eigene Forschungsdaten erhoben oder externe Datensätze verarbeitet haben. Allerdings bieten bislang nur wenige Zeitschriften effektive Möglichkeiten, die verwendeten Forschungsdaten im Kontext der zugehörigen Zeitschriftenartikel aufzurufen. Auch eine Zitation, Replikation oder eine Nachnutzung dieser Daten ist oftmals nicht möglich. Unser Paper beschreibt zunächst den Status Quo in Bezug auf das Forschungsdatenmanagement in wirtschaftswissenschaftlichen Fachzeitschriften. In einem zweiten Schritt werden die Ergebnisse einer Studie des DFG-geförderten Projekts EDaWaX (European Data Watch Extended) präsentiert und diskutiert, bei der die Richtlinien von Fachzeitschriften zur Übermittlung von Forschungsdaten untersucht wurden. Aufbauend auf den Ergebnissen der Studie wird die Frage aufgeworfen, wie eine sinnvolle Bereitstellung solcher publikationsbezogenen Forschungsdaten erfolgen kann.

### **Abstract**

Empirical studies to verify and refine theoretical models are increasingly important in economics. This is reflected in a rising number of empirical contributions to journals where the authors have collected their own research data or used external datasets. However, so far there have been few effective means to access the results of economic research within the framework of the corresponding article. Also to cite, to replicate or to make the research data that was used in a paper available for repurposing often is not possible. Initially our paper describes the status quo in regard to the management of research data in economic scholarly journals. In a second step the results of an empirical study conducted by the DFG-funded research project EDaWaX (European Data Watch Extended) are presented and discussed. The project evaluated the data availability policies of some economic scholarly journals. Based on the results of the study the question is raised how publication-related research data can reasonably be made available.

---

<sup>1</sup> Die in dieser Publikation gewählte männliche Form bezieht immer gleichermaßen alle Geschlechter mit ein. Auf eine Mehrfachbezeichnung wurde nur aufgrund einfacherer Lesbarkeit verzichtet.

## Einleitung

In den Wirtschaftswissenschaften existieren gegenwärtig schätzungsweise rund 8.000 bis 10.000 Fachzeitschriften<sup>2</sup>, davon werden nach heutigem Stand fast 500 Open Access bereitgestellt<sup>3</sup>. Der wissenschaftliche Zeitschriftenmarkt in den Wirtschaftswissenschaften wird daher nach wie vor von klassischen Subskriptionszeitschriften dominiert.

Längst nicht alle dieser vielen tausend Zeitschriften gehören zu den relevanten Publikationen der Fachdisziplin: So listet Thomson Reuters Journal Citation Reports (JCR)<sup>4</sup> im Jahr 2011 rund 320 wirtschaftswissenschaftliche Fachzeitschriften auf, die über einen Impact Factor verfügen. Das renommierte Handelsblatt Ranking vergab für das Jahr 2010 einen Impact Factor für 1250 Zeitschriften aus dem Bereich Volkswirtschaftslehre<sup>5</sup> und 750 Zeitschriften aus dem Bereich Betriebswirtschaftslehre<sup>6</sup>. Diese Zahlen zeigen, dass die Zahl der relevanten Zeitschriften noch einigermaßen übersichtlich ist.

Nicht alle diese Zeitschriften veröffentlichen empirische Beiträge. So gibt es eine nicht unerhebliche Anzahl von Zeitschriften die ausschließlich theoretische Arbeiten publizieren, oder Artikel deren Fokus auf Politikberatung liegt.

Insgesamt jedoch nimmt die Zahl an Veröffentlichungen in wirtschaftswissenschaftlichen Fachzeitschriften zu, bei denen Forscher eigene Datensätze erhoben und ausgewertet haben, oder bereits bestehende Datensätze für ihre Fragestellungen verwendet haben. Effektive Möglichkeiten die verwendeten Forschungsdaten im Kontext der zugehörigen Zeitschriftenartikel aufzurufen oder zu zitieren, bieten bislang allerdings nur wenige Journals. Auch eine Prüfung oder Nachnutzung dieser Daten zur Unterstützung des wissenschaftlichen Diskurses ist oftmals nicht möglich. Dabei ist eine Prüfung und Replikation der postulierten Forschungsergebnisse nicht nur vor dem Hintergrund zahlreicher Fälle von wissenschaftlichem Fehlverhalten<sup>7</sup> geboten, sondern es sprechen, wie

---

<sup>2</sup> [http://open-access.net/de/oa\\_in\\_verschiedenen\\_fachern/wirtschaftswissenschaften](http://open-access.net/de/oa_in_verschiedenen_fachern/wirtschaftswissenschaften) (Aufruf: 15. August 2012)

<sup>3</sup> Vgl. Directory of Open Access Journals (DOAJ): <http://www.doaj.org/doi?func=subject&cpid=17&uiLanguage=en> (Aufruf: 15. August 2012)

<sup>4</sup> [http://thomsonreuters.com/products\\_services/science/science\\_products/a-z/journal\\_citation\\_reports/](http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/)

<sup>5</sup> Handelsblatt Ranking VWL: <http://tool.handelsblatt.com/tabelle/?id=33> (Aufruf: 15. August 2012)

<sup>6</sup> Handelsblatt Ranking BWL: <http://tool.handelsblatt.com/tabelle/?id=34&so=1a&pc=25&po=0> (Aufruf: 15. August 2012)

<sup>7</sup> Ein bekanntes Beispiel für wissenschaftliches Fehlverhalten stellt der niederländische Sozialpsychologe Diederik Stapel, der zahlreiche seiner Forschungsdaten vorsätzlich gefälscht hat. Vgl. <http://www.nature.com/news/2011/111101/full/479015a.html>

beispielsweise der US-Ökonom B.D. McCullough argumentiert, verschiedene Gründe für die Replizierbarkeit von Forschungsdaten: „[...] *replication ensures that the method used to produce the results is known. Whether the results are correct or not is another matter, but unless everyone knows how the results were produced, their correctness cannot be assessed. Replicable research is subject to the scientific principle of verification; non-replicable research cannot be verified. Second, and more importantly, replicable research speeds scientific progress. [...] Third, researchers will have an incentive to avoid sloppiness. [...] Fourth, the incidence of fraud will decrease (McCullough, 2009,118f).*“

Dabei könnte von technologischer Seite heutzutage der Zugriff auf Forschungsdaten so leicht wie nie zuvor sei: Das Internet hat die vormals dominierenden Printpublikationen nun zunehmend mit Onlineausgaben der Zeitschriften ergänzt oder diese sogar ersetzt. Dadurch ist prinzipiell der Zugriff auf in Artikeln genutzte Forschungsdaten möglich und sogar deren weitere Bearbeitung und Nachnutzung. Damit solche Forschungsdaten aber verfügbar sind, müssen die Journals einerseits Richtlinien aufstellen, mit denen diese Daten eingefordert werden. Andererseits muss eine technische Infrastruktur für die Speicherung und das Auffinden von Forschungsdaten im Web bereitstehen.

Forschungsdatenrichtlinien manifestieren sich zumeist in so genannten Data (Availability) Policies, die Autoren dazu verpflichten, die genutzten Forschungsdaten zusätzlich zur Publikation bei den Herausgebern einer Fachzeitschrift einzureichen.<sup>8</sup> Gegenwärtig verfügt nur ein sehr geringer Anteil von Zeitschriften aus unterschiedlichen Gründen über solche Richtlinien. So gibt es auf Seiten der Herausgeber Befürchtungen, dass sich die Einführung einer solchen Policy negativ auswirken könnte: Geringere Einreichungsquoten werden hierbei ebenso genannt, wie potenziell zusätzlicher Aufwand durch komplexere Kommunikation mit Autoren einerseits, sowie durch Sichtung der Daten und eine Prüfung, ob die Vorgaben der Data Policy eingehalten werden andererseits.

---

<sup>8</sup> In der wissenschaftlichen Literatur wurde dieser Themenbereich vor allem durch B.D. McCullough et al (2006, 2007, 2008, 2009) bearbeitet. Weitere relevante Veröffentlichungen erschienen durch die Ökonomen Dewald, Thursby und Anderson (1986), sowie durch Anderson et al (2008) und Glandon (2010).



Auch der infrastrukturelle Aufwand durch eine Bereitstellung der übermittelten Daten im Internet wird als Hinderungsgrund angesehen.<sup>9</sup>

Infolgedessen konnte McCullough im Rahmen einer 2009 veröffentlichten Studie gerade einmal 14 solcher Zeitschriften in den Wirtschaftswissenschaften ausmachen, was ihn zu deutlicher Kritik veranlasste: "Results published in economic journals are accepted at face value and rarely subjected to the independent verification that is the cornerstone of the scientific method. Most results published in economics journals cannot be subjected to verification, even in principle, because authors typically are not required to make their data and code available for verification (McCullough/McGeary/Harrison, 2006, 1093f.)."

#### **Data Policies in wirtschaftswissenschaftlichen Fachzeitschriften**

Den aktuellen Status Quo in Bezug auf das Vorhandensein von Data Policies in wirtschaftswissenschaftlichen Fachzeitschriften zu evaluieren, war Ziel einer empirischen Untersuchung des DFG-geförderten Projekts „European Data Watch Extended“ (EDaWaX). In der Studie wurde untersucht, wie sich der Anteil von Zeitschriften mit entsprechenden Richtlinien in den letzten Jahren verändert hat, ob Zeitschriften mit Data Policies schwerpunktmäßig in bestimmten Verlagen oder Verlagsgruppen publiziert werden und wie Journals mit solchen Richtlinien in verschiedenen wirtschaftswissenschaftlichen Zeitschriftenrankings bewertet wurden. Ausführlich wurde ausgewertet, wie die Data Availability Policies der Zeitschriften jeweils im Detail ausgestaltet sind, welche konkreten Vorgaben den Autoren darin gemacht werden und in wie fern die Richtlinien eine Replizierbarkeit der publizierten Forschungsergebnisse ermöglichen.

#### Sample und Methodik

Für die Untersuchung wurde eine Sample aus insgesamt 141 Fachzeitschriften gebildet, das sich an einer Liste von Fachzeitschriften orientiert, die von den Ökonomen Bräuninger, Haucap und Muck (im Folgenden: BHM) im Jahre 2011 erstellt wurde. Diese Zeitschriften wurden seinerzeit durch Mitglieder der wirtschaftswissenschaftlichen Fachgesellschaft Verein für Socialpolitik (VfS)<sup>10</sup> hinsichtlich ihrer Relevanz und Reputation bewertet: „Die ausgewählten Zeitschriften

---

<sup>9</sup> Diese Einschätzungen werden auch von einer Onlinebefragung von fast 200 wirtschaftswissenschaftlichen Zeitschriften gestützt, die von den Autoren im Rahmen des EDaWaX-Projekts durchgeführt wurden. Eine Summary der Ergebnisse dieser Befragung wird im Herbst 2012 auf der Projektwebseite [www.edawax.de](http://www.edawax.de) publiziert.

<sup>10</sup> <http://www.socialpolitik.org/vfs.php?mode=start>

umfassen die wichtigsten in internationalen Ranglisten berücksichtigten Fachzeitschriften sowie einen Großteil der deutschsprachigen bzw. in Deutschland, Österreich oder der Schweiz herausgegebenen deutsch- und englischsprachigen VWL-Fachzeitschriften [...] (BHM, 2011, 2)“. Dieses Sample bot verschiedene Vorteile für die Auswertung: So ließen sich Zeitschriften aus dem deutschen Sprachraum mit den internationalen Top Journals vergleichen, und weitere Analysen in Bezug auf Relevanz und Reputation der Zeitschriften mit Data Availability Policy vornehmen.

Ergänzt durch vier Zeitschriften mit Data Availability Policies wurde insgesamt ein Sample genutzt, in dem solche Zeitschriften deutlich stärker vertreten sind, als dies durchschnittlich im Bereich der Wirtschaftswissenschaften der Fall sein dürfte, was jedoch aufgrund des Untersuchungsgegenstandes vertretbar ist. Zudem ist das Untersuchungssample durch Einbeziehung vieler Top-Journals überdurchschnittlich hoch gerankt.

Die 141 Zeitschriften unseres Samples wurden zunächst mittels Desktop-Research untersucht. Dabei wurden sowohl die Webseiten des Verlages als auch die der Herausgeber nach Richtlinien untersucht, die den Umgang mit Forschungsdaten regeln.<sup>11</sup> Für eine deskriptive Auswertung des Rankings der untersuchten Zeitschriften wurden Mittelwerte und Median für das Sample erstellt. Zudem wurden logistische Regressionen durchgeführt, um mögliche Zusammenhänge zwischen dem erzielten Ranking und dem Vorhandensein von Data Policies zu prüfen.

Daran anschließend wurden die Inhalte der aufgefundenen Data Policies anhand von Kriterien analysiert, die in der Literatur als wesentliche Elemente zur Ermöglichung von Replikationen benannt und durch eigene Überlegungen des Projekts ergänzt wurden.

Darüber hinaus wurden eine Onlinebefragung von Herausgebern von fast 200 Fachzeitschriften sowie zwei detaillierte Case Studies durchgeführt.

---

<sup>11</sup> Die Printausgaben der Zeitschriften wurden nicht geprüft.

### Deskriptive Ergebnisse

Mehr als 20% (29) der untersuchten 141 Zeitschriften verfügen über eine Data Policy, bei der Daten und weitere Materialien an die Herausgeber übermittelt werden müssen, die diese wiederum zumeist für Dritte bereitstellen. Sie entsprechen somit häufig dem von Harvard Professor Gary King formulierten „Replication Standard“ (King, 1995). Der relativ hohe Wert von über 20% ist aufgrund unserer Sampleauswahl aber keinesfalls als repräsentativ für die Wirtschaftswissenschaften anzusehen.

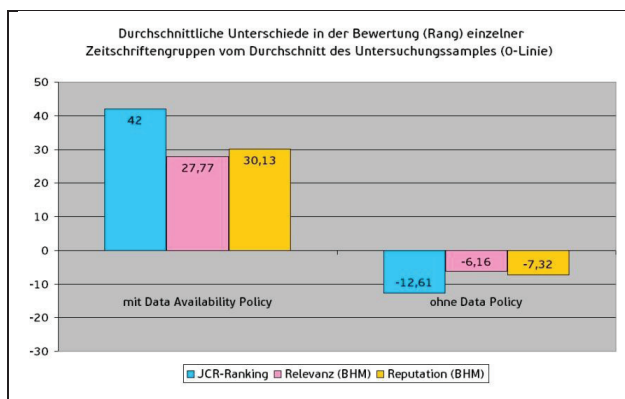
Zeitschriften mit solchen Richtlinien wurden am häufigsten durch die Verlage Wiley-Blackwell (6), und Elsevier (4) publiziert. Im Vergleich zeigte sich jedoch, dass anteilig vor allem Verlage aus dem universitären Kontext (z.B. Cambridge University Press) bzw. von wissenschaftlichen Fachgesellschaften (z.B. American Economic Association) über hohe bis sehr hohe Anteile an Zeitschriften mit Data Availability Policies verfügen.

Zudem konnten wir feststellen, dass Zeitschriften mit Data Availability Policy deutlich besser gerankt waren als Zeitschriften ohne solche Richtlinien:

Der Impact Factor dieser Zeitschriften im JCR liegt durchschnittlich fast 0,5 Punkte

höher als der Mittelwert unseres Samples. Beim Handelsblatt Ranking waren es durchschnittlich 0,21 Punkte mehr.

Ähnlich deutlich fallen die Ergebnisse bei der Platzierung der Zeitschriften mit Data Availability Policy im Vergleich zu solchen ohne



derartige Richtlinien aus: So wurden im JCR Zeitschriften mit Data Availability Policy durchschnittlich 42 Plätze höher bewertet als der Mittelwert. Bei den Rankings von BHM waren es durchschnittlich 27,8 Plätze (bei Relevanz) bzw. 30,1 Plätze (bei Reputation). Die durchgeführten Regressionsanalysen unterstützen diese Angaben, da für alle Rankings ein signifikanter mittlerer Zusammenhang zwischen einem

hohen Ranking und dem Vorhandensein einer Data Availability Policy ermittelt wurde.<sup>12</sup>

Somit ist festhalten, dass es vor allem die Top-gerankten Zeitschriften sind, die Data Availability Policies umsetzen. Dies mag einerseits mit den notwendigen zusätzlichen Ressourcen in Verbindung stehen, die für die organisatorische und technische Umsetzung einer solchen Richtlinie nötig sind. Andererseits kann vermutet werden, dass es gerade Publikationen in solchen prestigeträchtigen, hochgerankten Journals sind, die Wissenschaftlern die Anreize verschaffen, die oft umfangreiche und zeitaufwändige Dokumentation und Bereitstellung von Forschungsdaten vorzunehmen – eine Arbeit für die das Wissenschaftssystem bislang sonst keine angemessene Anerkennung liefert (Huschka et al, 2011).

#### Kriterien für die inhaltliche Untersuchung von Data Policies

Die Inhalte der aufgefundenen Data Availability Policies wurden in einem zweiten Schritt anhand von 9 inhaltlichen Kriterien analysiert, die in der Literatur als Wesentlich für das Gelingen von Replikationen charakterisiert werden und durch eigene Überlegungen ergänzt wurden. Folgende fachspezifische Empfehlungen für Data Policies wurden entwickelt:

- Die Richtlinien sollten verbindlich sein.
- Autoren sollten die verwendeten Datensätze an die Herausgeber übermitteln.
- Autoren sollten die verwendeten Berechnungscodes und Skripte an die Herausgeber übermitteln.
- Autoren sollten ausführliche Beschreibungen der übermittelten Daten sowie der Variablen und Merkmalsausprägungen der verwendeten Datensätze, am besten in Form eines „Data Dictionary“, an die Herausgeber übersenden.
- Wurden bei den Berechnungen (selbstgeschriebene) Programme verwendet (z.B. im Rahmen von Simulationen), sollten Autoren diese verwendeten Programme an die Herausgeber übermitteln.
- Die Datensätze und weitere Materialien sollten bereits vor der Publikation eines Artikels an die Herausgeber übermittelt werden.
- Wenn Ausnahmen von den Richtlinien (z.B. für Forschungen auf Basis von proprietären Daten) gestattet werden, sollte in den Richtlinien ein Vorgehen definiert sein, welche Daten dennoch an die Herausgeber übermittelt werden sollten (z.B. ausführliche Beschreibungen und Berechnungscodes).

---

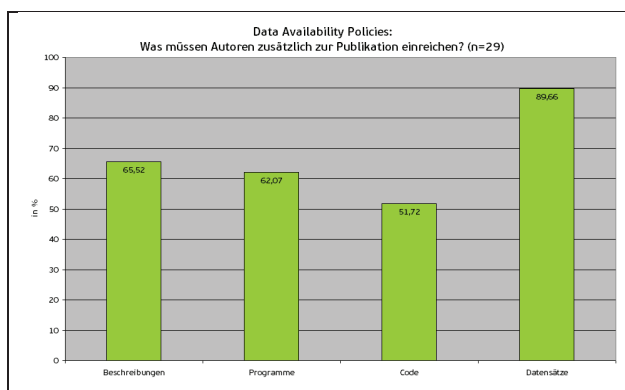
<sup>12</sup> Zwar erklärt die Data Availability Policy nur zu 8,7% bis 16% (je nach untersuchtem Ranking) das hohe Ranking von Zeitschriften mit einer Data Availability Policy, dies jedoch auf einem Signifikanzniveau zwischen 99,7% und 100%.

- Datensätze und weiteren Materialien sollten aus Gründen der Interoperabilität sowie der Langzeitarchivierung im ASCII-Format oder in anderen offenen Formaten an die Herausgeber übermittelt werden.
- Die Version der für die Berechnungen genutzten Software sowie des verwendeten Betriebssystems sollte bei der Übermittlung der Daten mit angegeben werden.

Unsere Prüfung der aufgefundenen Data Policies auf die oben genannten Empfehlungen ergab folgende Ergebnisse:

**Verbindlichkeit der Data Policies:** Wir konnten feststellen, dass 82,8% der Data Availability Policies für Autoren verpflichtend waren. Diese hohen Prozentsätze sind begrüßenswert, da vor allem verpflichtende Policies von den Autoren befolgt werden (McCullough et al., 2008).

**Übermittlung von Datensätzen:** Das Vorhandensein der genutzten Datensätze ist die erste Bedingung zur Durchführung von Replikationen. Fast 90% der untersuchten Data Policies sahen eine verpflichtende Übermittlung von Datensätzen vor. Unter den 10% der Zeitschriften, die keine verbindliche Übermittlung von Datensätzen vorsahen, waren einige, deren Fokus vor allem auf experimenteller Wirtschaftsforschung lag. Bei diesen ist die Übermittlung von Datensätzen zumeist optional.



**Übermittlung von Berechnungscode:** Replikationen sind ohne den passenden Berechnungscode sehr mühselig, da Forschende diesen erneut schreiben müssen. Ob dieser dann mit dem Berechnungscode des Autors übereinstimmt, ist

fraglich. Daher ermöglichen Zeitschriften, die ausschließlich die verwendeten Datensätze einfordern, nicht aber den dazu passenden Code, oftmals keine Replikationen (Dewald/Thursby/Anderson, 1986).

Unsere Untersuchung ergab, dass nur 51,7% der untersuchten Data Availability Policies die Übermittlung dieses Berechnungscodes einforderten. An dieser Stelle zeigen sich deutlich die qualitativen Probleme vieler untersuchten Richtlinien.

**Übermittlung von Beschreibungen:** Nur 65% der untersuchten Data Availability Policies forderten Beschreibungen zu den übermittelten Daten mit ein. Dieser Wert ist nicht zufriedenstellend, da eine Beschreibung der übermittelten Daten notwendig ist, um zu verstehen, welche der übermittelten Daten mit welchen Ergebnissen in einem Artikel korrespondieren (McCullough, 2007). Darüber hinaus ist die Beschreibung von Datensätzen, Variablen und Merkmalsausprägungen für die Durchführung von Replikationen sowie für die Nachvollziehbarkeit der Ergebnisse von hoher Relevanz.

**Übermittlung von Programmen:** 62% der untersuchten Data Availability Policies forderten die Übermittlung genutzter, meist selbst geschriebener Programme, wie sie häufig im Rahmen von Simulationen verwendet werden. Auch dieser Wert ist nicht zufriedenstellend, da es ohne die genutzten Programme oft nicht möglich ist, zu prüfen, ob die Ergebnisse korrekt berechnet wurden. Folglich können Ergebnisse dann nicht repliziert werden (McCullough/McGeary/Harrison, 2006).

**Zeitpunkt der Datenübermittlung:** Mehr als 90% der untersuchten Data Availability Policies sehen vor, dass Daten und Datensätze bereits vor der Veröffentlichung der Forschungspublikation an die Herausgeber übermittelt werden müssen. Hier haben die Herausgeber offensichtlich aus den Erfahrungen anderer Data Policies gelernt, die zeigen, dass eine verbindliche Übermittlung der Daten vor allem dann erfolgt, wenn es für die Wissenschaftler einen Anreiz gibt, die Daten zu publizieren. Nach der Publikation eines Artikels ist dieser Anreiz nicht mehr in dieser Form gegeben (Dewald/Thursby/Anderson, 1986).

**Definiertes Vorgehen im Fall proprietärer Daten:** Ausnahmen von den Richtlinien im Fall der Nutzung von proprietären oder vertraulichen Daten sehen mindestens 72,4% der Data Availability Policies vor, aber nur rund 60% verfügen über ein Verfahren, das regelt, welche Daten die Autoren dennoch an die Herausgeber einer Zeitschrift übermitteln müssen. Auch bei Nutzung proprietärer oder vertraulicher Daten sollten Autoren weiterhin dazu verpflichtet sein, Berechnungscode und

Beschreibungen zu den Daten weiter zu geben, so dass Replikationen im Prinzip auch bei solchen Forschungsdaten möglich sind.

**Formatvorgaben:** Vorgaben zur Nutzung bestimmter Formate, wie ASCII oder von offenen Formaten werden nur in Ausnahmefällen gemacht. Nur einzelne Zeitschriften gehen auf diese Problematik ein, und halten ihre Autoren an, die Daten (auch) in solchen Formaten bereitzustellen. Dabei sind offene Formate bzw. ASCII besonders unter den Gesichtspunkten der interoperablen Nutzung von Daten und Code, aber auch in Hinsicht auf eine Langzeitarchivierung der Daten sinnvoll (McCullough/McGeary/Harrison, 2006).

**Angabe der Version der genutzten Software und des Betriebssystems:** Die Angabe der genutzten Versionen der Software und des Betriebssystems ist wichtig, weil die erzielten Forschungsergebnisse nach genutzter Software und verwendetem Betriebssystem divergieren können (McCullough/Vinod, 2003).

Bis heute sieht keine der untersuchten Richtlinien eine Angabe des bei den Berechnungen verwendeten Betriebssystems vor und nur rund ein Drittel der Data Availability Policies fordert zudem die Angabe der genutzten Softwareversion.

## **Ergebnisse**

Zusammenfassend ist festzustellen, dass sich die Anzahl von Zeitschriften mit Data Availability Policy seit den letzten Forschungen von McCullough (2009) deutlich erhöht hat. Es ist somit ein gesteigertes Interesse der Herausgeber an solchen Richtlinien erkennbar.

Unsere Untersuchung hat ergeben, dass die analysierten Data Availability Policies teils erheblich in Umfang, Qualität und in der Granularität, in der Vorgaben für Autoren gemacht werden, divergieren. Dabei zeigte sich, dass die untersuchten Richtlinien nicht in allen Punkten inhaltlich gut ausgestaltet sind: So sind zwar rund 90% der Richtlinien verbindlich und sogar mehr als 90% verlangen die Übermittlung von Daten und Datensätzen bereits vor der Publikation des zugehörigen Artikels, jedoch fordern nur  $\frac{3}{4}$  der untersuchten Policies genutzte Programme oder Beschreibungen zu den Daten ein. Auch gewähren mehr als 70% der untersuchten Data Policies Ausnahmen von ihren Forschungsdatenrichtlinien im Fall der Verwendung proprietärer oder vertraulicher Daten, aber nur rund 60% haben

definiert, welche Daten und Angaben Autoren in solchen Fällen trotzdem übermitteln müssen. Hier zeigen sich deutliche inhaltliche Schwächen bei einigen Data Policies. Als problematisch muss insbesondere gelten, dass nur etwas mehr als die Hälfte der Richtlinien zugleich den Berechnungscode mit einfordert - für die angestrebte Replizierbarkeit von Forschungsergebnissen dürften solche Policies nicht ausreichen. Auch das weitgehende Fehlen von Metadaten zu den Versionen der Software und des Betriebssystems, welche bei der Berechnung der Ergebnisse genutzt wurde, ist kritikwürdig. Der weitgehende Verzicht auf Formatvorgaben ist speziell in Hinsicht auf eine interoperable Nutzung der Daten und die Langzeitverfügbarkeit von Forschungsdaten problematisch.

Aufgrund der erzielten Ergebnisse ist zu vermuten, dass nur etwa die Hälfte der untersuchten Data Availability Policies Replikationen gewährleisten dürfte.

Besonders gut ausgearbeitet war die Data Availability Policy der American Economic Review (AER)<sup>13</sup>. Dieser Richtlinie kann als Best Practice unter den aufgefundenen Data Policies bezeichnet werden. Damit, dass 34,5% der untersuchten Data Policies diese Richtlinie in identischer oder leicht modifizierter Form nutzen, ist eine positive Entwicklung in Richtung einer sich etablierenden Best-Practice in den Wirtschaftswissenschaften zu beobachten.

### **Infrastruktur zur Bereitstellung von publikationsbezogenen Forschungsdaten**

Das Vorhandensein von Data Availability Policies ist nur eine Bedingung für die Nachnutzung von Forschungsdaten oder zur Replikation der Forschungsergebnisse. Eine andere Bedingung ist, dass eine entsprechende technische Infrastruktur zur Bereitstellung dieser publikationsbezogenen Forschungsdaten zur Verfügung steht.

In der wissenschaftlichen Forschung wird dieses Thema vor allem von Institutionen bearbeitet, die die Bereitstellung von Forschungsdaten übernehmen (z.B. Wagner und Huschka, 2012; Huschka et al, 2011) als auch von übergeordneten Kommissionen und Projekten<sup>14</sup>, die teilweise auch dezidierte Vorschläge zur Entwicklung einer Forschungsdateninfrastruktur für publikationsbezogene Forschungsdaten enthalten.

---

<sup>13</sup> <http://www.aeaweb.org/aer/data.php/>

<sup>14</sup> Vgl. Allianz der deutschen Wissenschaftsorganisationen, 2010; Reilly et al, 2011; Kommission zur Zukunft der Informationsinfrastruktur, 2011; Wissenschaftsrat, 2011; Wissenschaftsrat, 2012



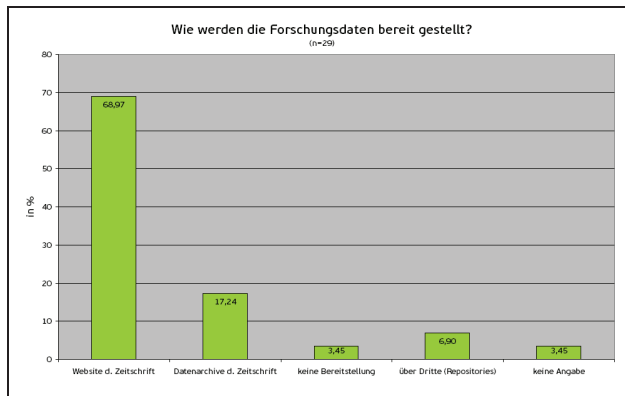
Auf Seiten der notwendigen technischen Infrastruktur für die Etablierung von zitierfähigen Forschungsdatenpublikationen gibt es sowohl in der Bundesrepublik als auch international übergreifende Ansätze, auf denen aufgebaut werden kann. So wurde im Jahr 2001 mit der Einrichtung der ersten vier Forschungsdatenzentren (FDZ) und zwei Datenservicezentren (DSZ) begonnen.<sup>15</sup> Ziel dieser Datenzentren ist es, die amtlichen Daten zur wissenschaftlichen Nutzung zur Verfügung zu stellen. Dies war bis dahin aufgrund von gesetzlichen Vorgaben nicht ohne weiteres möglich (Huschka et al, 2011). Neben diesen existieren weltweit unterschiedliche Datenarchive mit wirtschafts- und sozialwissenschaftlich relevanten Daten (z. B. Mikrodaten zu privaten Haushalten und Unternehmen oder Makrodaten zum Arbeitsmarkt oder zum Außenhandel). Diese Datenarchive verfügen über Erfahrungen bei der Beschreibung von Datensätzen und haben unter anderem den internationalen Metadatenstandard DDI<sup>16</sup> (Data Documentation Initiative) entwickelt. Ebenso verfügen viele Datenarchive über langjährige Erfahrungen im Bereich der Zugangsregelungen und Datenschutzbestimmungen bezüglich ihrer Daten (Reilly et al, 2011) sowie in Bezug auf den Langzeiterhalt dieser wichtigen Quellen von Forschung und Lehre (Mauer, 2011). Speziell solche Datenzentren sind daher auch als speichernde und hostende Institutionen eines publikationsbezogenen Datenarchivs in hohem Maße geeignet.

Im Zuge unserer Untersuchungen der Data Policies von wirtschaftswissenschaftlichen Journals evaluierten wir auch die eingesetzte technische Infrastruktur der Datenarchive von Fachzeitschriften. Dabei zeigte sich, dass eine Bereitstellung der Forschungsdaten zu fast 70% über die Webseiten der Zeitschriften erfolgt. Im Regelfall werden die Daten in einem zip-Container als „Supplementary Material“ zur Verfügung gestellt.

---

<sup>15</sup> Aktuell (Anfang August 2012) werden auf den Seiten des RatSWD 17 Forschungsdatenzentren und drei Servicedatenzentren nachgewiesen. <http://ratswd.de/dat/fdz.php>

<sup>16</sup> <http://www.ddialliance.org/>



Einige Zeitschriften nutzen zur Bereitstellung der Daten auch spezielle Webseiten („Data Archives“), auf der die Datensätze aller Ausgaben und Artikel einer Zeitschrift zum Download angeboten werden. Mit Ausnahme einer Übersicht, zu welchen Artikeln Forschungsdaten

existieren, bieten solche speziellen Webseiten aber keine weiteren Features.

Diese Praxis, Forschungsdaten in Form von zip-files in den Anhängen einer Publikation bereit zu stellen, kann als unzureichend angesehen werden, da die Daten nicht zitierbar und nur schwer auffindbar sind. Zudem existieren nur selten Metadaten zu den Forschungsdaten (und wenn, werden dafür keine standardisierten Metadatenschemata verwendet), was eine Nachnutzung der Daten und eine Replikation der Ergebnisse deutlich erschwert. Insbesondere wirkt sich hier negativ aus, dass es keine quodifizierten und von der Community akzeptierten Standards für Replikationen gibt, so dass die Qualität der Beschreibungen zu den Datensätzen teils erheblich variiert.

Etwas anders ist dies bei gut 17% der Fachzeitschriften mit Forschungsdatenrichtlinie, die für ihre Datenarchive spezielle Applikationen wie z.B. Dataverse<sup>17</sup> nutzen, die Funktionalitäten wie persistente Verlinkung und Zitierbarkeit der Forschungsdaten bieten. Über eine Anbindung an Forschungsdatenzentren verfügt hingegen keines der untersuchten Journals.

Um die Optionen von Infrastrukturlösungen für Journal Data Archives weiter zu untersuchen, werden im Rahmen des EDaWaX-Projekt derzeit fast 50 potenzielle Datenprovider dahingehend geprüft, ob diese ein publikationsbezogenes Forschungsdatenarchiv für die Wirtschaftswissenschaften hosten können, wie es gegenwärtig vom EDaWaX-Projekt entwickelt wird.

<sup>17</sup> Die Webseite von Dataverse ist verfügbar unter: <http://www.thedata.org>. Andere Softwarelösungen die ähnliche Funktionalitäten enthalten sind Nesstar (<http://www.nesstar.com>) oder TheDataHub (CKAN) (<http://www.thedatahub.org>).

Neben den FDZ und DSZ, den internationalen Datenarchiven wie ICPSR<sup>18</sup> und CESSDA-Verbund<sup>19</sup> wurden auch einzelne Bibliotheken, Bibliotheksverbünde und Archive in die Untersuchung mit einbezogen.

Im Rahmen einer Desktop Research wurde untersucht, welche dieser Institutionen auf die Anforderungen von EDaWaX passen. Speziell wurde geprüft, ob und gegebenenfalls in welcher Form diese Einrichtungen Speichermöglichkeiten für externe Forschungsdatensätze bieten, die für spezifische Publikationen tatsächlich genutzt wurden. Analysiert wurde ferner, welche Richtlinien für die Abgabe von externen Daten gelten, welche technische Infrastruktur dazu eingesetzt wird, welche Metadatenschemata Verwendung finden, ob und wenn wie Nutzer bei der Metadatengenerierung unterstützt werden. Auch die Frage nach ggf. anfallenden Gebühren für Hosting und Speicherung wurde in der Untersuchung berücksichtigt.

Die vorläufigen Ergebnisse der Desktop Recherche zeigten, dass von den untersuchten Institutionen nur eine Minderheit externe Daten annimmt. Zwar gaben 46% der untersuchten Datenzentren an, dass sie generell externe Datensätze annehmen, jedoch beschränkte sich dies oftmals nur auf Datensätze der eigenen Fachdisziplin. Mehr als 10% der untersuchten Datenzentren boten zudem ein Nutzerinterface an, mit dem Wissenschaftler Datensätze über NESSTAR einpflegen können.

Mit der Desktop Research sind die Arbeiten zu diesem Arbeitspaket jedoch noch nicht abgeschlossen. Im Herbst 2012 wird eine Onlinebefragung die Services der Datenzentren detailliert evaluieren.

### **Fazit:**

Im Zuge unserer Studie konnten wir feststellen, dass sich die Zahl der Zeitschriften mit Data Availability Policies in den Wirtschaftswissenschaften deutlich erhöht hat, wenngleich die Anzahl von 29 vor dem Hintergrund von tausenden solcher Zeitschriften eher dem viel zitierten „Tropfen auf den heißen Stein“ gleichkommt, als einem Paradigmenwechsel bei wirtschaftswissenschaftlichen Fachzeitschriften.

Die aufgefundenen Data Policies divergieren erheblich – sowohl hinsichtlich ihres Umfangs als auch hinsichtlich ihrer Qualität. So wird von vielen Richtlinien

---

<sup>18</sup> Interuniversity Consortium for Political and Social Research (ICPSR):  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

<sup>19</sup> Council of European Social Science Data Archives (CESSDA): <http://www.cessda.org/>

beispielsweise der für Replikationen wichtige Berechnungscode nicht mit eingefordert. Andere verzichten auf die Übermittlung von genutzten Programmen, oder verlangen keine dezidierten Beschreibungen der übermittelten Daten und ihrer Funktion. Vor diesem Hintergrund ist anzunehmen, dass ungefähr die Hälfte der untersuchten Richtlinien Replikationen nicht in vollem Umfang gewährleisten dürften. Auf der anderen Seite ist mit der Data Availability Policy der American Economic Review (AER) ein guter Standard geschaffen worden, durch den Replikationen zu einem hohen Prozentsatz erfolgreich verlaufen (Glandon, 2010). Damit, dass 34,5% der untersuchten Richtlinien diese Policy nachnutzen, ist ein Schritt in Richtung einer Best Practise unternommen worden.

Unsere Analysen zeigen aber auch, dass speziell die eingesetzte Infrastruktur zur Bereitstellung von publikationsbezogenen Forschungsdaten bei wirtschaftswissenschaftlichen Fachzeitschriften stark verbesserungswürdig ist. Gegenwärtig werden Forschungsdaten zumeist als zip-Datei an die Publikationen angehängt und sind über die Webseiten des Journals verfügbar. Da zumeist keine persistenten Identifikatoren vergeben werden und keine weiteren Metadaten zu den Daten hinzugefügt werden, sind die Forschungsdaten oft nicht zitier- und suchbar.

Zudem erfolgen auf den Servern der Herausgeber oder der Verlage keine Maßnahmen zum Datenmanagement oder zum Langzeiterhalt der Daten.

Eine wesentlich bessere Praxis wird von einer Minderheit von Journals betrieben, die Datenrepositories wie Dataverse für ihre Datenarchive nutzen. Hier finden sich sowohl persistente Identifikatoren als auch ein rudimentäres Set an Metadaten. Dennoch bieten auch diese publikationsbezogenen Forschungsdatenarchive keine Anbindung an größere Forschungsdatenzentren, die eine professionelle Datenvorhaltung gewährleisten könnten.

Im Rahmen des EDaWaX-Projekts werden wir in den nächsten Monaten hier ansetzen und das Potential einer Kooperation zwischen Fachzeitschriften und Forschungsdatenzentren im Rahmen von Umfragen und Interviews näher beleuchten.

#### **Literaturverzeichnis**

- Allianz der deutschen Wissenschaftsorganisationen: Grundsätze zum Umgang mit Forschungsdaten, Berlin 2010. Verfügbar unter: [http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_156.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_156.pdf)
- Anderson, R./Greene, W./McCullough, B.D./Vinod, H. D.: The Role of Data/Code Archives in the Future of Economic Research. In: Journal of Economic Methodology, Nr.1, S. 99-119.

- Bräuninger, Michael/Haucap, Justus/Muck, Johannes: Was lesen und schätzen Ökonomen im Jahr 2011?, DICE Ordnungspolitische Perspektiven, Nr. 18. Verfügbar unter: <https://econstor.eu/dspace/bitstream/10419/49023/1/667448497.pdf>
- Dewald, William G./Thursby, Jerry G./Anderson, Richard G.: Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. In: The American Economic Review, Vol. 76, No. 4 (1986), S. 587-603.
- Glandon, Phillip: Report on the American Economic Review Data Availability – Compliance Project, Vanderbilt University, November 2010. Verfügbar unter: [http://www.aeaweb.org/aer/-2011\\_Data\\_Compliance\\_Report.pdf](http://www.aeaweb.org/aer/-2011_Data_Compliance_Report.pdf)
- Huschka, Denis/Oellers, Claudia/Ott, Notburga/Wagner, Gert G.: Datenmanagement und Data Sharing: – Erfahrungen in den Sozial- und Wirtschaftswissenschaften. In: Büttner, Stephan / Hobohm, Hans-Christoph /Müller, Lars: Handbuch Forschungsdatenmanagement, Bad Honnef, 2011, S. 35 - 48. Verfügbar unter: <http://opus4.kobv.de/opus4-fhpotdam/files/208/HandbuchForschungsdatenmanagement.pdf>
- King, Gary: Replication, Replication. In: PS: Political Science and Politics, Nr. 28, 1995, S. 443-499. Verfügbar unter: <http://gking.harvard.edu/gking/files/replication.pdf>
- Mauer, Reiner: Das GESIS Datenarchiv für Sozialwissenschaften. Vortrag beim Workshop „Archivierung sozial- und wirtschaftswissenschaftlicher Datenbestände“ in der Deutschen Nationalbibliothek, Frankfurt, 15./16. September 2011. Verfügbar unter: [http://www.ratswd.de/ver/-docs\\_Archivierung\\_2011/mauer.pdf](http://www.ratswd.de/ver/-docs_Archivierung_2011/mauer.pdf)
- McCullough, B.D.: Got Replicability? The Journal of Money, Credit and Banking Archive. In: Econ Journal Watch, Nr. 4, 2007, S. 326-337.
- McCullough, B.D.: Open Access Economics Journals and the Market for Reproducible Economic Research. In: Economic Analysis and Policy, Nr. 39, 2009, S. 117-126.
- McCullough, B.D./McGeary, K. A. / Harrison, T.: Lessons from the JMCB Archive. In: Journal of Money, Credit, and Banking, Nr. 38, 2006, S. 1093-1107.
- McCullough, B.D./McGeary, Kerry Anne / Harrison, T.: Do economics journal archives promote replicable research? In: Canadian Journal of Economics, Nr. 41, 2008, S. 1406-1420.
- McCullough, B.D./McKittrick, Ross: Check the Numbers: The Case for Due Diligence in Policy Formation, Studies in Risk and Regulation, Fraser Institute, 2009. Verfügbar unter: <http://www.fraserinstitute.org/WorkArea/DownloadAsset.aspx?id=2486>
- McCullough, B.D./Vinod, H.D.: Verifying the Solution from a Nonlinear Solver: A Case Study." In: American Economic Review, Vol. 93, Issue 3, S. 873-892. DOI:10.1257/00028280322157133
- Reilly, S. / Schallier, W. / Schrimpf, S. / Smit, E. / Wilkinson, M.: Opportunities for Data Exchange: Report on Integration of Data and Publications, Oktober 2011. Verfügbar unter: <http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=ODE+Report+on+Integration+of+Data+and+Publications>
- Wagner, Gert G. / Huschka, Denis: Datenverfügbarkeit reicht nicht, um Replikationsstudien zur Routine zu machen. RatSWD Working Paper Series Nr. 194, 2012. Verfügbar unter: - [http://www.ratswd.de/download/RatSWD\\_WP\\_2012/RatSWD\\_WP\\_194.pdf](http://www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_194.pdf)
- Wissenschaftsrat: Empfehlungen zu Forschungsinfrastrukturen, Köln, 2011. Verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/10464-11.pdf>
- Wissenschaftsrat : Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, Berlin, 2012. Verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>

## **DOI® Registrierung sozial- und wirtschaftswissenschaftlicher Forschungsdaten mit da|ra**

Brigitte Hausstein<sup>1</sup>, Stefanie Grunow<sup>2</sup>

<sup>1</sup>Registration Agency für Social and Economic Data, GESIS

<sup>2</sup>Legal Affairs & Licensing, ZBW

### **Keywords:**

Forschungsdaten, Datenzitation, Persistent Identifier, DOI®-Namen

### **Zusammenfassung**

Im Rahmen von DataCite betreiben seit 2011 das GESIS Leibniz Institut für Sozialwissenschaften und die Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft die Datenregistrierungsagentur da|ra. Über die Vergabe von persistenten Identifiern in Form von DOI®-Namen werden Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften referenzier- und zitierbar gemacht. Dies entspricht auch den Zielen des internationalen Verbundes DataCite, Wissenschaftlern den Zugang zu Forschungsdaten zu erleichtern, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu steigern und somit die Einhaltung der Regeln guter wissenschaftlicher Praxis zu unterstützen.

In dem Beitrag wird sowohl die Struktur und Arbeitsweise von DataCite vorgestellt als auch die Einbindung des Projektes da|ra erläutert. Anschließend sollen die technische und organisatorische Umsetzung der DOI®-Registrierung sowie das da|ra Metadatenschema als Grundlage für das Informationssystem dargestellt werden. Das da|ra Metadatenschema ist ein speziell für die Registrierung von Forschungsdaten entwickeltes Metadatenschema. Es werden zudem die organisatorischen Rahmenbedingungen für die Registrierung über da|ra beleuchtet und die Policy, die die allgemeinen Registrierungsvoraussetzungen enthält, sowie das Service Level Agreement, welches die vertragliche Grundlage zwischen dem Publikationsagenten und da|ra zur Registrierung von Forschungsdaten bildet, vorgestellt.

### **Abstract**

Since 2011 the GESIS Leibniz-Institute for Social Sciences and the ZBW Leibniz Information Centre for Economics have been operating a DOI® registration service for social and economic research data. This service is offered in cooperation with DataCite, the international initiative to establish easier access to digital research data. This contribution introduces the structure and the workflows of DataCite and the way the Registration Agency da|ra is embedded in this organisation. Furthermore the paper describes the technical and organisational implementation of da|ra as well as its structural basis, the da|ra Metadata Schema. This schema is specially designed to

model metadata associated with social and economic research data. The organisational framework of the da|ra registration process is illustrated by introducing the da|ra Policy and the Service Level Agreement. While the Policy sets the general preconditions for the DOI<sup>®</sup> registration the Service Level Agreement serves as the contractual basis.

## 1. Einleitung

Während für Forschungspublikationen neben den traditionellen Angeboten der freie Zugang (*Open Access*) immer mehr zur gängigen Praxis wird, sind die Bemühungen hinsichtlich allgemein zugänglicher Datenpublikationen erst am Anfang. Obwohl grundsätzlich die Bereitschaft zur Weitergabe der Primärdaten existiert, scheitert dies oft an den fehlenden Kapazitäten, die für die Aufbereitung und Metadatenbeschreibung notwendig sind. Dies gilt auch für die Sozialwissenschaften, die im Vergleich zu anderen Disziplinen bereits eine ausgeprägte Kultur des „Data Sharings“ kennen.

Die Verbreitung der Forschungsergebnisse erfolgt fast ausschließlich noch über Publikationen in Fachzeitschriften. Die „Allianz der deutschen Wissenschaftsorganisationen“ hat jedoch Ende Juni 2010 in den „Grundsätze(n) zum Umgang mit Forschungsdaten“<sup>1</sup> eine Regelung für Primärdaten gefordert, um bei Wissenschaftlerinnen und Wissenschaftlern das Bewusstsein für den Handlungsbedarf und für den Nutzen von Primärdaten-Infrastrukturen zu schärfen. Von Seiten der Forschungsfinanzierer wird zunehmend gefordert, nicht nur die Forschungspublikationen sondern auch die entstandenen Primärdaten im Sinne von *Good Scientific Practice* öffentlich zugänglich zu machen. Daraus ergibt sich die besondere Bedeutung einer reinen Datenpublikation, mit allen Möglichkeiten der eindeutigen Identifikation und kompakten Zitierung, die für Textpublikationen bereits Standard sind.

Darüber hinaus zeigt sich in allen Wissenschaftsbereichen auch die Dringlichkeit der Kopplung von Datenarchivierung und wissenschaftlicher Literaturpublikation.

Die Trennung von Forschungspublikation und zugrundeliegenden Daten erschwert die Evaluation der Publikation und schränkt die Nachvollziehbarkeit der dargestellten Ergebnisse ein. Bislang erfolgt die Verbindung von Daten- und

---

<sup>1</sup> [www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/](http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/)

Forschungspublikation nur punktuell. Voraussetzung für die Verbindung von Forschungsprimärdaten und wissenschaftlicher Publikation ist jedoch neben der Langzeitarchivierung der Daten und entsprechender Qualitätssicherung, die Möglichkeit zur Publikation von Daten mit eindeutiger Identifizierbarkeit und Referenzierbarkeit.

## 2. Persistent Identifier Systeme

Ein Weg zur Lösung der geschilderten Problematik ist der Einsatz von speziellen Persistent Identifier. Deren Funktion entspricht in etwa einer ISBN-Nummer bei gedruckten Werken, die lediglich ein einziges Mal vergeben wird. Hinzu kommt die Unterscheidung zwischen dem Identifier und der Lokation eines Objekts, die es ermöglicht, das Objekt unabhängig von seinem Speicherort zu identifizieren. Dies unterscheidet die Persistent Identifier von einer *Universal Resource Locator (URL)*. Zur Sicherstellung der eindeutigen Vergabe und der Zuweisung von Kennung und Speicherort bedarf es eines automatisierten Dienstes. Jedem Persistent Identifier werden dabei Adressinformationen, z.B. ein URL zugewiesen. Von zentraler Bedeutung sind hier geeignete organisatorische Maßnahmen, die Verweise auf die tatsächlichen Speicherorte der Ressourcen aktuell halten. Programme können dann über einen sogenannten Resolverdienst den zitierten Persistent Identifier auflösen, so dass ein Zugang zu den zitierten Forschungsdaten möglich wird.

Es existieren mittlerweile für die Identifikation von elektronischen Textpublikationen diverse Systeme von Persistent Identifier, die technisch die Basis für einen Service auch zur Identifizierung von Daten leisten können: *Archival Research Key (ARK)*, *Digital Object Identifier (DOI®)*, *Handle*, *Library of Congress Control Number (LCCN)*, *Life Science Identifiers (LSID)*, *Persistent URL (PURL)*, *Uniform Resource Name (URN)* und weitere. Auf einen gemeinsamen Standard haben sich die verschiedenen Nutzergemeinden jedoch noch nicht geeinigt, da die Systeme im Prinzip gut ineinander überführbar sind. Um die langfristige Eignung zu beurteilen ist hier weniger die technische als die organisatorische Ausgestaltung relevant.



## 2.1 Das DOI®-System

Das DOI®-System<sup>2</sup> wurde von der *Association of American Publishers* entwickelt und wird gegenwärtig von der *International DOI® Foundation (IDF)* verwaltet. Die IDF besteht seit 1998 und unterstützt die Rechteverwaltung für geistiges Eigentum in digitalen Netzwerken indem sie die Entwicklung und Verbreitung des DOI®-Systems als gemeinsame Infrastruktur für das *Content Management* fördert. Die IDF ist als *not-for-profit* Organisation registriert und wird von einem *Executive Board*, das von den Mitgliedern des IDFs gewählt wird, kontrolliert. Die Mitgliedschaft ist offen für alle Organisationen, die sich mit elektronischem Publizieren und den damit verbundenen Technologien beschäftigen.

Das DOI®-System ist ein verwaltetes System für die persistente Identifikation von Inhalten, die in digitalen Netzwerken angeboten werden. Es kann für die Identifizierung von physikalischen, digitalen oder anderen Objekten benutzt werden. Die Identifikatoren (DOI®-Namen) führen direkt zum Speicherort des bezeichneten Objektes. Technisch basiert das DOI®-System auf der von der Corporation for National Research Initiatives® (CNRI)<sup>3</sup> entwickelten *Handle Technology*<sup>4</sup>. Es wird ergänzt durch ein Metadatenmodell, um die zum Objekt gehörenden Metadaten mit dem DOI®-Namen zu verbinden. Auf der Basis der gemeinsamen Policy und technischen Infrastruktur der IDF wird das DOI®-System durch einen Zusammenschluss von Registrierungsagenturen umgesetzt. Dieses Lizenzmodell wird fälschlicher Weise mit einer kommerziellen Ausrichtung des DOI®-Systems verwechselt. Jede Registrierungsagentur kann jedoch über ihr eigenes Businessmodell für die Vergabe der DOI®-Namen entscheiden. Zum Beispiel bieten GESIS und ZBW den Service da|ra kostenfrei an (vgl. Abschnitt 3.1).

Das DOI®-System hat gute Aussichten auf Verbreitung und Langlebigkeit. Dies wird nicht zuletzt durch verbindliche Verträge zwischen Registrierungsagentur und Nutzern erreicht, die eine längere Zeitperspektive versprechen. Insbesondere die internationalen wissenschaftlichen Fachverlage setzen fast durchgängig auf die Verwendung von DOI®-Namen (vgl. Brase et al. 2009). Das DOI®-System überzeugt daher auch für den Einsatz im Wissenschaftsbereich.

---

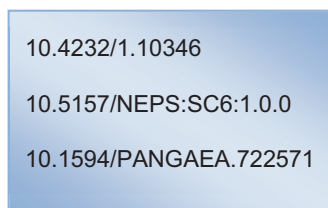
<sup>2</sup> [www.doi.org](http://www.doi.org)

<sup>3</sup> [www.cnri.reston.va.us](http://www.cnri.reston.va.us)

<sup>4</sup> [www.handle.net](http://www.handle.net)

## 2.2 Struktur und Resolving eines DOI®-Namens

Ein DOI®-Name besteht genau wie ein *Handle* immer aus einem Präfix und einem Suffix, wobei beide durch einen Schrägstrich getrennt werden und das Präfix stets mit „10.“ beginnt (vgl. Abbildung 1). Das Präfix, das beispielsweise einem bestimmten Datenzentrum zugeordnet ist, ermöglicht die Bildung einer unbegrenzten Anzahl von DOI®-Namen, indem auf der Basis des Präfixes und verschiedener Suffixe eine beliebige Reihe von eindeutigen Identifiern gebildet werden können.



10.4232/1.10346  
10.5157/NEPS:SC6:1.0.0  
10.1594/PANGAEA.722571

Abb. 1: Beispiele für DOI®-Namen

Um einen DOI®-Namen zur zugehörigen URL aufzulösen gibt es verschiedene Möglichkeiten, die alle auf dem zentral betriebenen Handle Server<sup>5</sup> basieren. Zum einen kann er über das vom CNRI angebotene Resolver-plugin<sup>6</sup> eingegeben und aktiviert werden. Eine andere Möglichkeit ist die Verwendung des Proxy Servers des DOI®-Systems (<http://dx.doi.org/>) bzw. des Handle Systems (<http://hdl.handle.net/>). Die Eingabe des DOI®-Namens zusammen mit der vorangestellten Zeichenkette <http://dx.doi.org/> in den Eingabeschlitz jedes beliebigen Browsers führt den Nutzer direkt zum Speicherort des Objektes bzw. zu einer Webseite, die dieses und die Zugangsbedingungen ausführlich beschreibt.

## 2.3 DataCite

DataCite<sup>7</sup> ist ein 2009 in London gegründetes internationales Konsortium mit inzwischen zwanzig Mitgliedern aus zwölf Ländern. Während DataCite eine bei der IDF akkreditierte DOI®-Registrierungsagentur ist, fungieren GESIS und ZBW als Vollmitglieder in DataCite als DOI® Allocation Agency und weisen daher selbst DOI®-Namen an Publikationsagenten zu. (Vgl. Abbildung 2)

<sup>5</sup> Dieser wird ab 2012 nicht mehr vom CNRI sondern von der AT&T, einer für die UNO in Genf tätigen Gesellschaft betreut.

<sup>6</sup> [www.handle.net/resolver](http://www.handle.net/resolver)

<sup>7</sup> [www.datacite.org](http://www.datacite.org)

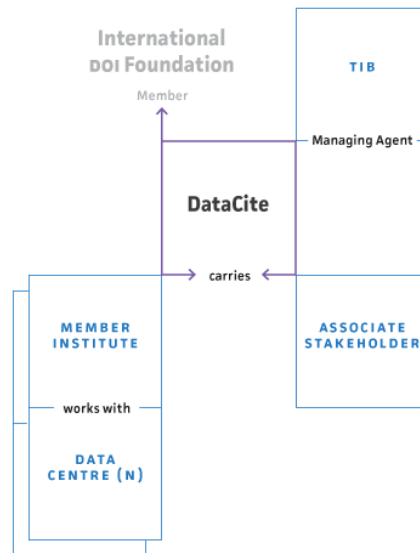


Abb. 2 Organisationsstruktur von DataCite

Unter den zwanzig Mitgliedern befinden sich vier assoziierte Mitglieder, die zwar nicht selbst als Allocation Agency tätig sind, jedoch ein großes Interesse an dem Ziel aller DataCite-Mitglieder haben, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu fördern und dieses gemeinsam im Konsortium verfolgen möchten. So wurde in diesem Rahmen im Juni 2012 eine gemeinsame Erklärung von DataCite und der International Association of Scientific, Technical & Medical Publishers veröffentlicht, die datenhaltende Organisationen (z.B. Datenzentren, Bibliotheken, Forschungsinstitute) und Verlage dazu ermutigen soll, wissenschaftliche Artikel und die ihnen zugrundeliegenden Datensätze zu verlinken.<sup>8</sup>

DataCite bietet neben der Mitgliedschaft in der IDF eine international abgestimmte Vorgehensweise bei technischen Lösungen, Standards, *Best Practises* und Workflows, sowie ein gemeinsames Metadatenschema und einen *MetaDataStore*, in dem die von den DataCite Mitgliedern registrierten DOIs<sup>®</sup> mit den dazugehörigen Metadaten nachgewiesen werden.<sup>9</sup>

---

<sup>8</sup> [www.datacite.org/node/65](http://www.datacite.org/node/65)

<sup>9</sup> [mds.datacite.org](http://mds.datacite.org)

Seit 2010 ist GESIS Mitglied in DataCite. Die ZBW wurde 2011 als viertes deutsches Mitglied<sup>10</sup> in DataCite aufgenommen. Mit der Mitgliedschaft in DataCite erhalten GESIS und ZBW die rechtliche Möglichkeit, DOIs® zu vergeben und schufen durch ihren Beitritt zu dem Konsortium die Voraussetzung für die Etablierung eines Registrierungsservices für Forschungsdaten auf der Basis von DOI®-Namen.

### 3. Registrierungsagentur da|ra

GESIS hat im Frühjahr 2010 ein Pilotprojekt unter dem Titel da|ra<sup>11</sup> (Registrierungsagentur für sozialwissenschaftliche Daten) zur Etablierung eines Registrierungssystems für Forschungsdaten der Sozialforschung im deutschsprachigen Raum gestartet. Ziel war es, eine Infrastruktur zu entwickeln, die es ermöglicht, Forschungsdatenbestände mit DOI®-Namen zu versehen und sie mit ihren Titeln, Themen, Autoren, Provenienzen, Methoden und Zugangsmöglichkeiten so umfassend wie möglich nachzuweisen sowie findbar und zitierbar zu machen.

In Kooperation mit DataCite wurde dazu mit der technischen Implementierung eines Registrierungstools begonnen und ein spezifisches über das von DataCite hinausgehende Metadatenmodell entwickelt. Begonnen wurde mit der Registrierung der Forschungsdaten des GESIS-Datenarchivs um die technischen und organisatorischen Lösungen zu testen.

In einer zweiten Projektphase (bis Ende 2012) wird der Service weiteren Forschungsdatenzentren, Datenservicezentren und institutionalisierten Forschungsprojekten angeboten. Schwerpunkt in dieser Phase ist aber auch die Entwicklung des DOI®-Registrierungsservices für Wirtschaftsdaten auf der Basis der vorhandenen technischen Lösung. Dazu wurde die Zusammenarbeit mit der ZBW aufgenommen. Dies lag nahe, da die Nutzergemeinschaften beider Infrastruktureinrichtungen aus eng verwandten Disziplinen stammen.

Die Einbeziehung weiterer Datenrepositorien und Datenkuratoren als Publikationsagenten ist für die Ausbauphase (ab 2013) geplant. In dieser Phase wird der Dienst in Abhängigkeit von der Entwicklung eines *Self-Archiving Services* bei GESIS auch Einzeldatenproduzenten aus dem sozialwissenschaftlichen Bereich angeboten. Über ein von der DFG gefördertes Projekt werden zusätzlich die Entwicklung von

---

<sup>10</sup> Weitere Mitglieder aus Deutschland sind die Technische Informationsbibliothek (TIB) Hannover sowie die Deutsche Zentralbibliothek für Medizin (ZB MED) in Köln.

<sup>11</sup> [www.gesis.org/dara](http://www.gesis.org/dara)

Standards und Guidelines für die DOI<sup>®</sup>-Registrierung sowie die Entwicklung eines erweiterten Nachweissystems realisiert.

### 3.1 Organisation und Verwaltung der Registrierung

Die DOI<sup>®</sup>-Registrierung erfolgt bei da|ra in enger Kooperation mit den datenhaltenden Organisationen, den so genannten Publikationsagenten, die für die Pflege und Speicherung der Forschungsdaten sowie für die Metadatenpflege zuständig sind. Die Datensätze verbleiben bei den Datenzentren, da|ra speichert die Metadaten und macht die registrierten Inhalte über eine eigene Datenbank recherchierbar.

Um dies zu gewährleisten, sind die Rahmenbedingungen und Voraussetzungen für die DOI<sup>®</sup>-Registrierung in einer Policy festgehalten. Daneben hat die Policy eine erläuternde Funktion, um den Publikationsagenten den Hintergrund und den Ablauf einer DOI<sup>®</sup>-Registrierung mit da|ra darzulegen. Insbesondere beinhaltet die Policy die Anforderungen, denen ein Objekt genügen muss, um einen DOI<sup>®</sup>-Namen zu erhalten. Zum Beispiel müssen Metadaten für jedes Objekt vorhanden sein und die Persistenz des Objektes sichergestellt sein. Des Weiteren beschreibt das Dokument die wesentlichen Verantwortlichkeiten, die bei der DOI<sup>®</sup>-Registrierung auf Seiten der Publikationsagenten und der Betreiber von da|ra entstehen.

Über die allgemeine Policy hinausgehend, werden der Workflow und sämtliche Verantwortlichkeiten im Registrierungsprozess in einem Service Level Agreement verbindlich vereinbart. Das Service Level Agreement ist ein bilateralen Vertrag zwischen dem Publikationsagenten und der Betreibereinrichtung von da|ra, die fachlich für den betreffenden Publikationsagenten zuständig ist. Das Service Level Agreement beinhaltet Fragen zur Qualitäts- und Persistenzsicherung (Daten und Metadaten), zur Rechtesituation an den Datensätzen und Metadaten, Versionierungen, zur Verfügbarkeit des Services sowie dessen Funktionalitäten. Um einen optimalen Beginn des Registrierungsservices für den jeweiligen Publikationsagenten zu gewährleisten, ist es notwendig, sich bereits vor Abschluss des Service Level Agreements über wichtige Rahmenbedingungen der Registrierung zu verständigen und diese im Vertrag festzuhalten.

Dazu gehört es, festzulegen, auf welchem Weg die Metadaten des Publikationsagenten an da|ra übermittelt werden (z.B. über die webbasierte Erfassungsmaske oder einen xml-Import), auf welcher Granularitätsstufe die Objekte registriert werden (z.B. Studienebene, Mikrodaten, usw.) und welche Struktur der DOI®-Name haben soll. Bei der Festlegung der Struktur des DOI®-Namens ist die am Anfang stehende "10." als Kennzeichen für das DOI®-System unveränderbar. Das Präfix wird von zentral von DataCite und CNRI dem Publikationsagenten zu gewiesen. Das Suffix ist frei gestaltbar, sollte jedoch eine einheitliche Struktur haben. Dieses bietet auch die Möglichkeit, eine Versionierung der Datensätze innerhalb des DOI®-Namens widerzuspiegeln. Das Suffix muss innerhalb eines Präfixes immer eineindeutig sein.

Sowohl bei der stetigen Weiterentwicklung der Policy durch die da|ra-Betreiber als auch bei der Festlegung der Details im Service Level Agreement werden in Zusammenarbeit mit dem Publikationsagenten die Best Practice Empfehlungen von DataCite berücksichtigt.

Während der Etablierungsphase von da|ra wird der Registrierungsservice kostenneutral angeboten. Nach Abschluss der Einführungsphase werden die Betreiber von da|ra prüfen, ob die Erhebung einer Gebühr erforderlich ist. Die Publikationsagenten haben im Falle der Einführung eines Kostenmodells durch da|ra die Möglichkeit, sich aus der vertraglichen Vereinbarung zu lösen und ihr Präfix zu behalten, auch wenn sie nach Ende der Beziehung zu da|ra von einer anderen registrierenden Einrichtung betreut werden. Ziel ist es jedoch, den Service so zu gestalten, dass die Betreiberkosten auch langfristig auf einem niedrigen Niveau gehalten werden können und die Nutzung des Basisservices weiterhin kostenfrei bleibt.

### 3.2 Technische Implementation

Die technische Implementation des Registrierungstools erfolgt über eine serviceorientierte Architektur. Zentrale Komponenten, wie beispielsweise das Registrieren eines DOI®-Namens oder das Indexieren, wurden in separate Services ausgelagert. Entsprechend verfügt auch das zentrale Informationssystem über Schnittstellen, die als Webservice angesprochen werden können.

Die Metadaten werden in einer Datenbank gemäß DDI Standard abgespeichert. Dadurch werden ein Im- und Export der Daten im XML-Format auf einfache Weise ermöglicht und ein Austausch mit den Publikationsagenten direkt unterstützt. Für die Funktionalität der Suche wird ein Indexierungsserver eingesetzt (SOLR). Die Oberfläche zum Editieren der Metadaten wird von Projektmitarbeitern oder den Publikationsagenten benutzt um die Metadaten zu pflegen. Die Edition wird durch die Verwendung von kontrollierten Vokabularen unterstützt, damit möglichst standardisierte Metadaten erstellt werden.

### 3.3 Das Metadatenmodell

Neben dem technischen DOI®-Registrierungsservice leistet da|ra auch die Übernahme von Metadatenbeschreibungen in sein Datenbanksystem. Damit wird eine differenzierte Suche aller registrierten Datensätze ermöglicht, eine inhaltliche Beschreibung zur Verfügung gestellt und die Voraussetzung zur einheitlichen Zitation der Daten geschaffen. Unter Berücksichtigung der besonderen Herausforderungen der Dokumentation digitaler Datenobjekte wurde für das Informationssystem ein eigenes Beschreibungsschema (vgl. Hausstein et al. 2011) entwickelt. Dieses basiert auf dem Metadatenschema des GESIS-Datenbestandskatalogs und wurde in Anlehnung an das DataCite-Metadatenschema<sup>12</sup> erweitert. Durch das Mapping mit dem DDI Standard<sup>13</sup> und durch die Verwendung kontrollierter Vokabulare ist die Interoperabilität der erfassten Ressourcen mit anderen Datenbeständen auch im internationalen Rahmen gewährleistet.

Das aktuelle da|ra Metadatenschema (Version 2.2.1) beinhaltet neun obligatorische Felder (Kernelemente), die für die Registrierung von Daten verpflichtend sind. Fünf dieser Pflichtelemente müssen vom Publikationsagenten bereitgestellt werden: „Typ der Ressource“, „Titel“, „Primärforscher“, „URL“, „Veröffentlichungsdatum“, „Verfügbarkeit“. Die drei weiteren Kernelemente „DOI®“, „Publikationsagent“ und „Registrierungsagentur“ werden von da|ra intern erzeugt. Diese Pflichtfelder sind im DataCite-Metadatenschema ebenfalls *mandatory elements*. Des Weiteren umfasst das Schema 24 optionale Elemente, die zur detaillierten Beschreibung der Daten verwendet werden können. Vier administrative Elemente werden systemintern geliefert.

---

<sup>12</sup> Dieses basiert auf ISO 690-2, dem Dublin Core Metadata Element Set, Version 1.1 und dem DOI-Metadaten Kernel der IDF und liegt z.Z. in der Version 2.2 vor: <http://schema.datacite.org/>.

<sup>13</sup> <http://www.ddialliance.org>

Ein wesentliches Element der Datenbeschreibung ist die Möglichkeit der Versionierung, die gerade im Bereich von Primärdaten für die eindeutige Referenzierung eines Datensatzes unerlässlich ist. Da jede Version mit einem eigenen DOI®-Namen und eigenen Metadaten versehen ist, wird die eindeutige Bezugnahme auf den einer Analyse zugrundeliegenden Datensatz und damit z.B. auch die Verifizierung von Analyseergebnissen ermöglicht.

### 3.4 da|ra Services

Die Registrierungsagentur da|ra bietet eine vollständige Infrastruktur für die DOI®-Registrierung und die Metadatenverwaltung. Die Betreiber von da|ra sehen sich als Dienstleister für Daten- und Forschungszentren, die ihre Primärdaten mit DOI®-Namen registrieren wollen. Dabei kann es sich um Umfragedaten, Aggregatdaten, Microdaten aber auch um qualitative Daten handeln.

Im Rahmen des Registrierungsprozesses erhält jeder Datensatz nach Übermittlung der erforderlichen Metadaten einen eindeutigen DOI®-Namen, wobei Granularität und Namensgestaltung vom Publikationsagenten festgelegt werden können. Der DOI®-Name kann dann sofort über einen Resolverdienst zur entsprechenden URL aufgelöst werden. Das Serviceangebot von da|ra umfasst gegenwärtig die DOI®-Namensregistrierung, eine DOI®-Resolvermöglichkeit, ein umfangreiches Metadatenbeschreibungsschema, das Metadatenhandling in einer Datenbank sowie verschiedene Möglichkeiten zur Übertragung der Metadaten. Darüber hinaus können interessierte Nutzer über eine einfache und erweiterte Suche in den Metadatenbeständen recherchieren.

Gegenwärtig nutzen acht sozialwissenschaftliche Forschungsdatenzentren und Datenarchive den da|ra Service. Diese haben insgesamt ca. 5700 Studien/Datensätze registriert, wobei der größte Anteil daran noch beim GESIS Datenarchiv liegt. Zusätzlich wurden ca. 4400 Metadatensätze von der *iLibrary* der OECD, die ihre Bestände über die Partnerschaft mit der DOI®-Registrierungsagentur *crossref* registriert haben, in das Informationssystem übernommen. Somit stehen mehr als 10.000 Metadatensätze für Recherchen im da|ra Informationssystem zur Verfügung.



## Ausblick

Im Rahmen der Etablierungsphase des Projektes haben GESIS und ZBW an der Anpassung des Registrierungssystems gearbeitet, um die Veränderungen zu berücksichtigen, die durch die technischen und organisatorischen Neuerungen bei

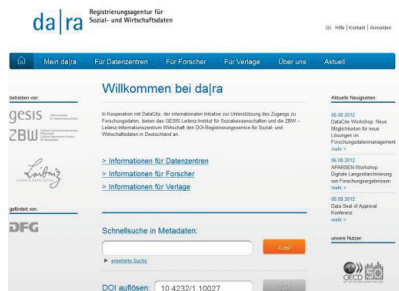


Abb. 3: da|ra Services

DataCite sowie durch die Erweiterung des Anwendungsbereichs von da|ra entstanden sind. Das betraf die Überarbeitung des Metadatenschemas, die entsprechende Anpassung der Datenbank, den Anschluss des Registrierungstools an den neugestalteten *MetaDataStore* von DataCite

sowie das Update der da|ra Policy und des Service Level Agreements. Die webbasierte Eingabemaske wurde umgestaltet und eine facettierte Suche sowie ein *help desk* für die Publikationsagenten implementiert. Das überarbeitete Angebot von da|ra ist seit dem 15. September 2012 unter der neuen Domain [www.da-ra.de](http://www.da-ra.de) online verfügbar. Besuchen Sie uns auf unserer Website!



## Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. 24. Juni 2010.  
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze/>
- Askitas, N. (2010): What Makes Persistent Identifiers Persistent? Working Paper No. 147. RatSWD Working Paper Series June 2010.  
[http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_147.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_147.pdf)
- Arbeitsgruppe Fachinformation (2009): Rahmenkonzept für die Fachinformation. Vorlage zur Sitzung des Ausschusses der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) am 29.09.2009.
- Bellini, E., Cirinnà, C. et al. (2008): Persistent Identifiers distributed system for Cultural Heritage digital objects [http://www.bl.uk/ipres2008/presentations\\_day2/38\\_Lunghi.pdf](http://www.bl.uk/ipres2008/presentations_day2/38_Lunghi.pdf)
- Bleuel, J. (2000): Zitation von Internet-Quellen (Citing of Internet sources). In: Hug, T. (Hrsg.): Wie kommt Wissenschaft zu Wissen? Band1: Einführung in das wissenschaftliche Arbeiten. Hohengehren: Schneider Verlag.
- Blue Ribbon Task Force (2010): Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)
- Brase, J. (2010): DataCite - A global registration agency for research data Working Paper No 149. RatSWD Working Paper Series July 2010.
- Brase J. et al. 2009: Approach for a global joint registration agency for research data. Information Services & Use. In: Neuroth H. et al. (Hrsg.): nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
- Brase, J. & Klump, J. (2007): Zitierfähige Datensätze. Primärdaten-Management durch DOIs. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. [http://dc110dmz.gfz-potsdam.de/content/std-doi/upload/pdf/Brase\\_Wisskomm\\_2007.pdf](http://dc110dmz.gfz-potsdam.de/content/std-doi/upload/pdf/Brase_Wisskomm_2007.pdf)
- DFG (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“; Denkschrift der Deutschen Forschungsgemeinschaft.  
[http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_pra\\_xis\\_0198.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_pra_xis_0198.pdf)
- Diepenbroek, M. & Grobe, H. (2007): PANGAEA® als vernetztes Verlags- und Bibliothekssystem für wissenschaftliche Daten. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. [http://www.fz-juelich.de/zb/datapool/page/1000/Diepenbroek\\_Abstract.pdf](http://www.fz-juelich.de/zb/datapool/page/1000/Diepenbroek_Abstract.pdf)
- Dittert, N., Diepenbroek, M. & Grobe, H. (2002): Data and information management for the CMTT synthesis. Manuskript <http://epic.awi.de/Publications/Dit2002b.pdf>
- GESIS Report 4/10. Der Aktuelle Informationsdienst für die Sozialwissenschaften. GESIS - Leibniz Institute für Sozialwissenschaften Mannheim September 2010.  
[http://www.gesis.org/fileadmin/upload/institut/presse/gesis\\_report/gesis\\_report\\_10\\_04.pdf](http://www.gesis.org/fileadmin/upload/institut/presse/gesis_report/gesis_report_10_04.pdf)
- Hausstein, B. (2012): Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten. Serviceleistungen der Registrierungsagentur daJra. RatSWD Working Paper Series 193.  
[http://www.ratswd.de/download/RatSWD\\_WP\\_2012/RatSWD\\_WP\\_193.pdf](http://www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_193.pdf)
- Hausstein, B. & Zenk-Möltgen, W. (2011): daJra – Ein Service der GESIS für die Zitation sozialwissenschaftlicher Daten. In: Schomberg, S., Leggewie, C. & Puschmann, C. (Hrsg.) Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland. 20/21. September 2011 Köln. Beiträge der Tagung.  
[www.hbz-nrw.de/Tagung\\_Digitale\\_Wissenschaft.pdf](http://www.hbz-nrw.de/Tagung_Digitale_Wissenschaft.pdf)
- Hausstein, B., Zenk-Möltgen, W., Wilde, A. & Schleinstein, N.: daJra Metadatenschema. Version 1.0 GESIS-Working Papers 2011|14. GESIS – Leibniz-Institut für Sozialwissenschaften 2011.  
 DOI:10.4232/10.mdsdoc.1.0
- Klump, J., Bertelmann, R. et al. (2006): Data publication in the open access initiative. Data Science Journal, Volume 5, Number 79.

Neuroth, H., Oßwald, A., Scheffel, R. Strathmann, S. Huth, K. (2007): nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. nestor – Network of Expertise in Long-Term Storage of Digital Resources. Göttingen. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>

Metadata for the Publication and Citation of Scientific Primary Data (Version 3.0) [http://www.icdp-online.org/contenido/std-doi/upload/pdf/STD\\_metadata\\_kernel\\_v3.pdf](http://www.icdp-online.org/contenido/std-doi/upload/pdf/STD_metadata_kernel_v3.pdf)

Paskin, N. (2000): Digital Object Identifier: implementing a standard digital identifier as the key to effective digital rights management. The International DOI Foundation Kidlington, Oxfordshire, United Kingdom [http://www.doi.org/doi\\_presentations/aprilpaper.pdf](http://www.doi.org/doi_presentations/aprilpaper.pdf)

Parsons, M. A., Duerr, R. & Minster, J.-B.: Data Citation and Peer Review. In: Eos, Transactions, American Geophysical Union, Vol. 91, No. 34, 24 August 2010, PAGES 297–304  
<http://aurora.gmu.edu/spaceweather/images/2010EO340001.pdf>

Riding the Wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. October 2010.  
<http://goo.gl/WrxO>

Reilly, S. (2010): Digital Object Repository Server: A Component of the Digital Object Architecture. D-Lib Magazine January/February 2010. Volume 16, Number 1/2  
<http://www.dlib.org/dlib/january10/reilly/01reilly.print.html>

Uhlir, P. F. (2003): Discussion Framework. In Esanu, J. M. & Uhlir, P. F. (Hrsg.): The Role of Scientific and Technical Data and Information in the Public Domain. Washington DC: The National Academies Press.

Uhlir, P. F. & Schroder, P. (2008): Chapter 8 - Open Data for Global Science. In Fitzgerald, B.: Legal Framework for e-Research: Realising the Potential. Sydney: Sydney University Press Law Books 39 (2008) 188. <http://www.austlii.edu.au/au/journals/SydUPLawBk/2008/39.html>

## Internet

Allianz der deutschen Wissenschaftsorganisationen  
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>

California Digital Library: ARK <http://www.cdlib.org/inside/diglib/ark/>

Corporation for National Research Initiatives® (CNRI) <http://www.cnri.reston.va.us>

crossref <http://www.crossref.org>

da|ra <http://www.gesis.org/dara> | <http://www.da-ra.de>

DataCite <http://www.datacite.org>

<http://www.datacite.org/whatisdatacite>

<http://www.datacite.org/node/65>

<http://mds.datacite.org>

Data Documentation Initiative <http://www.ddalliance.org/>

Digital Object Identifier (DOI®) System <http://www.doi.org/>

GESIS - Leibniz Institut für Sozialwissenschaften (Mannheim, Köln, Berlin) <http://www.gesis.org>

Homepage der Technischen Informationsbibliothek (TIB) Hannover zu eigener DOI-Registrierungsagentur <http://www.tib-hannover.de/de/die-tib/doi-registrierungsagentur/>

Homepage der Technischen Informationsbibliothek (TIB) Hannover zu Projekt CODATA  
<http://www.tib-hannover.de/de/die-tib/projekte/codata/>

Library of Congress: Structure of the LC Control Number [http://www.loc.gov/marc/lcn\\_structure.html](http://www.loc.gov/marc/lcn_structure.html)

Life Sciences Identifiers (LSID) <http://lsids.sourceforge.net/>

OECD iLibrary <http://www.oecd-ilibrary.org>

Persistent Uniform Resource Locator (PURL) <http://purl.oclc.org/docs/index.html>

Projekt Nationales Bildungspanel (NEPS)

Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten [http://www.std-doi.de/front\\_content.php](http://www.std-doi.de/front_content.php)

The Handle System® <http://www.handle.net>

Uniform Resource Name (URN) Syntax <http://tools.ietf.org/html/rfc2141>

ZBW <http://www.zbw.eu>

---

## **re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositoryn. Ein Werkstattbericht**

Heinz Pampel<sup>1</sup>, Hans-Jürgen Goebelbecker<sup>2</sup>, Paul Vierkant<sup>3</sup>

<sup>1</sup> Deutsches GeoForschungsZentrum GFZ, Potsdam

<sup>2</sup> Karlsruher Institut für Technologie (KIT)

<sup>3</sup> Humboldt-Universität zu Berlin

### **Zusammenfassung**

Um Wissenschaftler beim „data sharing“ zu unterstützen, bedarf es u.a. Infrastrukturen mit denen Forschungsdaten zugänglich und nachnutzbar gemacht werden können. Diese Infrastrukturen werden vermehrt unter dem Begriff Forschungsdaten-Repository gefasst. Die Landschaft dieser Repositoryn ist heterogen und unübersichtlich. Die Initiative re3data.org hat das Ziel diese Repositoryn in einem web-basierten Verzeichnis zu erschließen und so eine Orientierung über bestehende Datensammlungen zu bieten. Der Beitrag beschreibt Hintergrund und Anliegen des Projektes.

### **Abstract**

Researchers need support in data sharing, e.g. via infrastructures that allow access and reuse of research data. These infrastructures can be summarised under the term Research Data Repository. The landscape of these repositories is heterogeneous and complex. The aim of re3data.org is to create a web-based registry of Research Data Repositories. The initiative re3data.org will index Research Data Repositories to provide an overview on existing research data collections. The paper describes the background and purpose of the project.

### **1. Einführung**

Die Diskussion um die Zugänglichkeit von Forschungsdaten gewinnt an Relevanz. Die Forderungen internationaler Wissenschaftsorganisationen nach dem „offenen Zugang zu wissenschaftlichem Wissen“ (Berliner Erklärung, 2003) aus dem Jahr 2003 und die von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung 2007 veröffentlichten „Principles and Guidelines for Access to Research Data from Public Funding“ (OECD, 2007) sind zwei von vielen Referenzen einer breiten Debatte, die vielfältige Akteure im System Wissenschaft beschäftigt. Motor dieser Diskussion ist die Digitalisierung von Forschung und Lehre. Die Forderung nach „langfristige[r] Sicherung und [...] grundsätzlich[em] offenen Zugang zu Daten aus öffentlich geförderter Forschung“, die die deutschen Wissenschaftsorganisationen 2010 in „Grundsätze zum Umgang mit Forschungsdaten (Allianz 2010) formuliert haben, kann als Mission der digitalen Wissenschaft

verstanden werden. Ziel ist es, die „Produktivität und Wettbewerbsfähigkeit der Wissenschaft“ (Allianz 2010) zu fördern, indem die Nachprüfbarkeit (z.B. im Rahmen der Qualitätssicherung) und die Nachnutzung (z.B. zur Vermeidung von Doppelerhebungen der Daten) ermöglicht werden.

Eine anhaltende Herausforderung stellt eine interdisziplinär geltende Definition des Begriffes Forschungsdaten dar. Im Folgenden werden unter diesem Begriff digitale Daten gefasst, die je nach Disziplin in unterschiedlichsten Formen und Formaten vorliegen. Diese können während eines Forschungsprozesses entstanden sein oder Grundlage dieses Prozesses sein.<sup>1</sup>

Verstärkt wirken sogenannte „Data Policies“ auf Wissenschaftler und deren Umgang mit Forschungsdaten.<sup>2</sup> Hervorzuheben sind die empfehlenden und verpflichtenden Anforderungen von Förderorganisationen und wissenschaftlichen Zeitschriften, die Mittelempfänger und Autoren auffordern, die Zugänglichkeit der Daten, die im Rahmen eines Projektes entstehen oder Grundlage einer Publikation sind, sicher zu stellen. So sind z. B. Antragsteller der National Science Foundation (NSF) durch eine „Data Sharing Policy“ gefordert, „to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.“ (NSF, 2011a) Weiter fordert die NSF, dass Maßnahmen zur Umsetzung dieser Richtlinie in einem „Data Management Plan“ spezifiziert werden (NSF, 2011b). Auch die Deutsche Forschungsgemeinschaft (DFG) erwartet seit 2010 im Rahmen ihres Programms „Sachbeihilfe“ Aussagen zu Verfahren, „um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen“. Dabei sollen „sofern vorhanden [...] Angebote bestehender Datenrepositorien“ berücksichtigt werden (DFG, 2012). Dieses Anliegen wird vermehrt auch in den „Editorial Policies“ von wissenschaftlichen Zeitschriften deutlich. Beispielhaft seien hier die Anforderungen der Nature-Zeitschriftenfamilie genannt. In diesen heißt es: „authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications in material transfer agreements“. Angeregt wird die Zugänglichmachung der Daten „via public repositories.“ (Nature, 2011).

---

<sup>1</sup> Diese Definition wurde abgeleitet von Kindling (2009) und Kommission Zukunft der Informationsinfrastruktur (2011).

<sup>2</sup> Siehe dazu ausführlich: Pampel & Bertelmann (2011).

Vor dem Hintergrund der skizzierten Entwicklungen manifestieren sich Strategien zur Zugänglichmachung von Forschungsdaten. Kategorisierungen dieser Publikationsstrategien finden sich bei Dallmeier-Tiessen (2011) und Lawrence et al. (2011). Angelehnt an Dallmeier-Tiessen können drei Strategien der Zugänglichmachung unterschieden werden: 1. Die Veröffentlichung von Forschungsdaten als eigenständiges Informationsobjekt in einem Repositorium. 2. Die Veröffentlichung von Forschungsdaten mit textueller Dokumentation, als sogenanntes „Data Paper“<sup>3</sup>. 3. Die Veröffentlichung von Forschungsdaten als Anreicherung einer interpretativen Text-Publikation („Enriched Publication“<sup>4</sup>). Gemein ist diesen Publikationsstrategien, dass es einer Infrastruktur bedarf, die die Speicherung und Zugänglichmachung der Daten möglichst dauerhaft sicherstellt. Diese Infrastrukturen werden im Folgenden unter dem Begriff Forschungsdaten-Repositorien gefasst.

## 2. Forschungsdaten-Repositorien

Die Europäische Kommission stellte 2009 fest: „The landscape of data repositories across Europe is fairly heterogeneous, but there is a solid basis to develop a coherent strategy to overcome the fragmentation and enable research communities to better manage, use, share and preserve data.“ (European Commission, 2009). Das Zitat beschreibt die heutige unübersichtliche Landschaft von Infrastrukturen, die Forschungsdaten speichern und zugänglich machen. Forschungsdaten-Repositorien sind durch die Daten geprägt, die sie speichern. Anders als Text-Repositorien, die mehrheitlich Informationsobjekte, die im PDF-Format gespeichert sind, zugänglich machen und sich im Kontext von Open Access etabliert haben, speichern Forschungsdaten-Repositorien vielfältige Dateiformate.<sup>5</sup>

Bisher haben sich nur wenige Untersuchungen der weltweiten Landschaft von Forschungsdaten-Repositorien angenommen: Marcial & Hemminger beschreiben in einer 2010 veröffentlichten Studie die Charakteristik von 100 Repositorien, einen

---

<sup>3</sup> Siehe hierzu Chavan & Penev (2011).

<sup>4</sup> Eine Beschreibung des Konzeptes der „Enriched Publication“ findet sich bei Woutersen-Windhouwer & Brandsma (2009).

<sup>5</sup> Während sich Anfang der 2000er Jahre die Open Archives Initiative (OAI) formierte, die die Standardisierung und Vernetzung von Repositorien, die textuelle Informationsobjekte (wie Qualifikationsarbeiten und Pre- und Post-Prints) im Sinne des Open Access zugänglich machen, förderte, steht eine vergleichbare Entwicklung im Bereich der Forschungsdaten-Repositorien noch aus.



ähnlichen Ansatz verfolgt Schaaf 2011. Erhellend ist der Blick auf disziplinäre Untersuchungen. Eine eindrucksvolle Landschaft an Daten-Repositorien bietet die Biomedizin. Die 2012-Ausgabe der „Molecular Biology Database Collection“ des biomedizinischen Journals „Nucleic Acids Research“ weist 1.380 Infrastrukturen nach, auf denen biomedizinische Forschungsdaten gespeichert werden können (Galperin & Fernández-Suárez, 2012). Im Rahmen des Projektes European Life Science Infrastructure for Biological Information (ELIXIR) wurden 200 dieser Infrastrukturen näher untersucht. Diese 200 Repositorien werden von 100 Institutionen betrieben. Insgesamt sind 350 Personen in den Betrieb der Repositorien involviert, die durch mehrere hunderttausend Wissenschaftler pro Jahr etwa 60 Millionen Zugriffe verzeichnen. Die jährlichen Betriebskosten für diese 200 Repositorien liegen bei etwa 30 Millionen Euro (ELIXIR, n.d.). Um den nachhaltigen Betrieb dieser Repositorien-Landschaft zu garantieren, wurde ELIXIR in das „European Strategy Forum on Research Infrastructures (ESFIR)“ aufgenommen. Anliegen von ESFIR ist die strategische Förderung von Forschungsinfrastrukturen, die für die Wettbewerbsfähigkeit des Europäischen Forschungsraums (ERA) von zentraler Bedeutung sind. Bereits im ESFRI-Einsatzantrag des Rats der Europäischen Union wird deutlich, dass Forschungsinfrastrukturen heute keinesfalls nur physische Infrastrukturen wie Forschungsschiffe oder Beschleuniger sind, sondern auch digitale Infrastrukturen, wie z.B. „electronic archiving systems for scientific publications and databases“ (ESFRI, 2006).

Eine grobe Systematisierung dieser Repositorien in institutionelle, disziplinäre, multidisziplinäre und projektspezifische Infrastrukturen ist möglich. Beispiel für ein institutionelles Forschungsdaten-Repositorium ist Edinburgh DataShare.<sup>6</sup> Dieses „online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh“ (Edinburgh DataShare, o.J.) wird mit der Software DSpace realisiert und wurde im Zeitraum 2007-2009 aufgebaut (Rice, 2009). Im Juni 2012 sind 19 Datensätze auf dem Repositorium gespeichert.

Ein weiteres Beispiel für ein solches institutionelles Daten-Repositorium ist Open Data LMU<sup>7</sup>, dieses 2010 gestartete Repositorium „steht [...] allen Wissen-

---

<sup>6</sup> <http://datashare.is.ed.ac.uk>

<sup>7</sup> <http://data.ub.uni-muenchen.de>

schaftler/innen der LMU<sup>8</sup> als Publikationsplattform für Primärdaten“ zur Verfügung (Schallehn, 2010). Im Juni 2012 sind 34 Datensätze auf dem Repository gespeichert. Verwendet wird die Software EPrints.

Im Bereich der disziplinären Forschungsdaten-Repositories sind Beispiele wie GenBank oder PANGAEA - Data Publisher for Earth & Environmental Science<sup>9</sup> herauszuheben. GenBank<sup>10</sup>, wurde 1982 ins Leben gerufen<sup>11</sup> und definiert sich als „public database of nucleotide sequences and supporting bibliographic and biological annotation.“ Betrieben wird diese Infrastruktur, die Nukleotidsequenzen von mehr als 250.000 Spezies nachweist, vom National Center for Biotechnology Information (NCBI) (Benson et al., 2012). PANGAEA versteht sich als „Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research.“ (PANGAEA, n.d.). Dieses Repository wird vom Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI) und vom Zentrum für Marine Umweltwissenschaften der Universität Bremen (MARUM) betrieben. Aufgebaut wurde PANGAEA im Rahmen des Projektes „Paläoklima-Datenzentrum“, das vom Bundesministerium für Bildung und Forschung (BMBF) von 1994 bis 1997 gefördert wurde (Diepenbroek et al., 1999). 2011 speichert PANGAEA über eine halbe Million geowissenschaftlicher Datensätze (Schindler, Diepenbroek, & Grobe, 2012).

Einige Infrastrukturen verfolgen auch einen multidisziplinären Ansatz, z.B. FigShare<sup>12</sup>, eine 2011 gestartete „permanent research data storage and sharing platform“ der Verlagsgruppe Macmillan, oder LabArchives<sup>13</sup>, eine mit dem Verlag BioMed Central kooperierende privatwirtschaftlich organisierte „platform for publishing scientific data“ (Hrynaskiewicz, 2012).

Vielfältig ist auch die Landschaft der projektspezifischen Repositorien. Exemplarisch kann hier die vom Deutsches GeoForschungsZentrum GFZ betriebene Scientific Drilling Database (SDDDB)<sup>14</sup> genannt werden, die Bohrdaten, die im Rahmen des Scientific Continental Drilling Program (ICDP) entstehen, offen zugänglich macht (Klump & Conze, 2007). Ein weiteres Beispiel ist das Repository des Bern Digital

---

<sup>8</sup> Ludwig-Maximilians-Universität München

<sup>9</sup> <http://www.pangaea.de>

<sup>10</sup> <http://www.ncbi.nlm.nih.gov/genbank>

<sup>11</sup> Zur Geschichte von GenBank siehe: Cravedi (2008).

<sup>12</sup> <http://figshare.com>

<sup>13</sup> <http://www.labarchives.com>

<sup>14</sup> <http://www.scientificdrilling.org>

Pantheon Project<sup>15</sup>, in dem hochauflösende Fotografien und maßstabsgetreue Visualisierungen des Pantheon in Rom frei zugänglich gemacht wurden. Darüber hinaus gibt es viele weitere Typen von Forschungsdaten-Repositoryn, die teils nur schwer zu charakterisieren sind.

Betrachtet man die zukünftige Bedeutung dieser Infrastrukturen für eine digital arbeitende Wissenschaft, muss eine mangelnde Forschung und Entwicklung konstatiert werden. Der wissenschaftspolitische Wille einer offenen Wissenschaft (Kroes, 2012a; Kroes, 2012b) ist ohne verlässliche sowie vertrauenswürdige und damit auch dauerhafte Infrastrukturen, die Wissenschaftler beim „sharing“ ihrer Daten unterstützen, zum Scheitern verurteilt. Befragungen von Betreibern von Forschungsdaten-Repositoryn zeigen, dass insbesondere die finanzielle Absicherung dieser Infrastrukturen für längere Zeiträume (über 5 Jahre) unsicher ist (Kuipers & Van der Hoeven, 2009).

### **3. re3data-org: Ein Werkstattbericht**

Im Auftrag der Europäischen Kommission wurde 2010 eine Vision des Umgangs mit Forschungsdaten im Jahr 2030 beschrieben. Nach dieser sollen Wissenschaftler in Zukunft in der Lage sein, „to find, access and process the data they need.“ Weiter sollen Forschende, die Daten erheben, diese zukünftig in verlässlichen und vertrauenswürdigen Repositoryn speichern, die auf Basis internationaler Standards arbeiten (High Level Expert Group on Scientific Data, 2010). Mit Blick auf die heterogene Landschaft der Forschungsdaten-Repositoryn ist die Umsetzung dieser Vision eine zentrale Herausforderung für die digital arbeitende Wissenschaft.

Vor dem Hintergrund der skizzierten Forderung nach „data sharing“, die in der jüngsten Vergangenheit vermehrt auf die wissenschaftliche Praxis wirkt (siehe Abschnitt 1), und der heterogenen Landschaft der Forschungsdaten-Repositoryn (siehe Abschnitt 2) verfolgt das Projekt re3data.org - Registry of Research Data Repositories<sup>16</sup> den Aufbau eines Verzeichnisses von Forschungsdaten-Repositoryn.

---

<sup>15</sup> <http://www.digitalpantheon.ch/repository>

<sup>16</sup> <http://re3data.org>

Die Initiative hat die Erfassung und den Nachweis von Forschungsdaten-Repositorien aller Disziplinen in einem zentralen web-basierten Recherche- und Nachweissystem zum Ziel.

Das Projekt wird im Zeitraum 2012-2014 von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Projektpartner sind die Abteilung Bibliothek und Informationsdienste des Deutschen GeoForschungsZentrums GFZ<sup>17</sup>, das Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin<sup>18</sup>, sowie die Bibliothek des Karlsruher Instituts für Technologie (KIT)<sup>19</sup>. Vernetzungspartner ist die Deutsche Initiative für Netzwerkinformation (DINI)<sup>20</sup>, in deren Kontext die drei Projektpartner seit längerem rund um den Betrieb von Repositorien zusammenarbeiten und 2009 ein Positionspapier zum Thema Forschungsdaten veröffentlicht haben (Dallmeier-Tiessen et al., 2009).

Zentrales Anliegen des Projektes ist es, Wissenschaftlern eine Orientierung in der heterogenen Landschaft der Forschungsdaten-Repositorien zu geben. Dabei wird sowohl die Rolle der Erhebenden, die z.B. durch Förderorganisationen oder Zeitschriften aufgefordert sind Daten zugänglich machen, als auch die Rolle der Nutzenden, z.B. Forschende, die Daten von Dritten nachnutzen möchten, fokussiert. Weiter soll Förderorganisationen und infrastrukturellen Dienstleistern, wie Datenzentren, Rechenzentren und Bibliotheken, eine Übersicht über die Landschaft der Forschungsdaten-Repositorien gegeben werden. Darüber hinaus wird das Ziel verfolgt, einen Beitrag zur Schaffung eines „ecosystem of data repositories“ (Van der Graaf & Waaijers, 2011) zu leisten. Das Verzeichnis soll auch den Entwicklungsstand der Repositorien abbilden und kann so z. B. auch bei der Identifikation von Disziplinen helfen, in denen es eine lückenhafte Repositorien-Landschaft gibt. Übergeordnetes Ziel von re3data.org ist es, einen Beitrag zur Förderung des „data sharings“ in der Wissenschaft zu leisten.

In der ersten Phase des Projektes steht, ausgehend von der konzeptionellen Entwicklung des Verzeichnisses, die Recherche und Erschließung bestehender Forschungsdaten-Repositorien, sowie die technische Realisierung des Verzeichnisses im Fokus.

---

<sup>17</sup> <http://www.gfz-potsdam.de>

<sup>18</sup> <http://www.ibi.hu-berlin.de>

<sup>19</sup> <http://www.bibliothek.kit.edu>

<sup>20</sup> <http://www.dini.de>

Während bestehende Listen von Forschungsdaten-Repositories mehrheitlich nur die Disziplin und den Betreiber eines Repositoriums auszeichnen<sup>21</sup>, ist re3data.org bestrebt, eine möglichst umfassende Beschreibung eines Repositoriums zu bieten und dabei z.B. die unterschiedlichen Zugangsbedingungen zu dem Dienst und den darin vorgehaltenen Forschungsdaten darzustellen. Dieses Anliegen einer umfassenden Darstellung von Forschungsdaten-Repositories prägt die Konzeption des Verzeichnisses und wirkt auf dessen technische Realisierung.

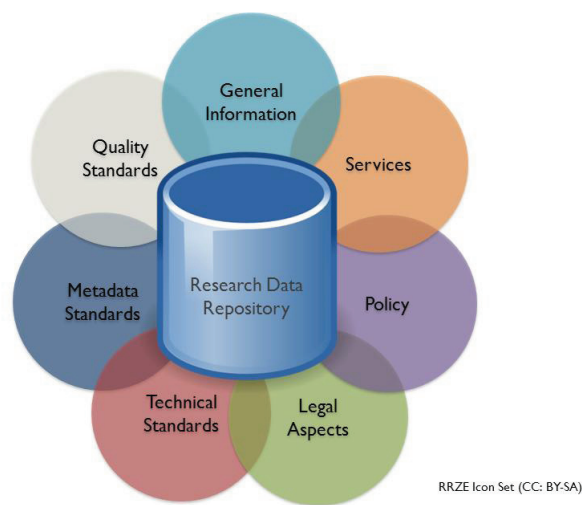


Abbildung 1: Aspekte eines Repositoriums, die im Vokabular (Version 1.0) von re3data.org betrachtet werden.

Ausgehend von einer Erhebung von rund 400 Daten-Repositories wurde ein Vokabular zur Beschreibung von Forschungsdaten-Repositories auf Basis einer Untersuchung, nach zuvor entwickelten Kriterien, von 20 zufällig gewählten Repositorien entwickelt (Vierkant et al., 2012).

Bei der Erhebung wurde auf bestehende Sammlungen zurückgegriffen, die durch eigene Recherchen ergänzt wurden. Weiter wurde das Projekt, nach der Kommunikation des Projektvorhabens im April 2012, von verschiedenen Repositorien-Betreibern über ggf. zu erschließende Dienste informiert.

---

<sup>21</sup> Siehe z.B. die auf der Website von DataCite integrierte Liste von Repositorien (<http://datacite.org/repolist>), oder die im Open Access Directory dokumentierten Repositorien ([http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)).

Das Vokabular gliedert sich in seiner Version 1.0 in folgende Bereiche: Allgemeine Informationen, Verantwortlichkeiten, Policies, rechtliche Aspekte, technische Standards, Metadaten-Standards, Qualitätsstandards und Services (siehe Abb. 1). Dank des Vokabular ist es zum Beispiel möglich in re3data.org abzufragen, welches zertifizierte Forschungsdaten-Repositorium einem Datenproduzenten die Speicherung von geowissenschaftlichen Daten unter einer Creative-Commons-Lizenz erlaubt.

Im Kontext der Qualitätsstandards wurden bestehende Anforderungen an Repositorien im Kontext der Zertifizierung eruiert (u.a. Braun et al., 2010; CCSSS, 2011; Data Seal of Approval, 2010; ESF & EUROHORCs, 2011; ICSU WDS, 2011).<sup>22</sup> Um der heterogenen Landschaft und den vielfach fehlenden Standards im Bereich der Forschungsdaten-Repositorien zu entsprechen, sind die formulierten Anforderungen an Repositorien, die in dem Verzeichnis indexiert werden, niedrigschwellig gehalten. Neben Pflichtangaben bei der Erschließung der Forschungsdaten-Repositorien sind Angaben zu Zugang<sup>23</sup> und Lizenzierung der Forschungsdaten Mindestanforderungen, die Repositorien-Betreiber erfüllen müssen, um in re3data.org aufgenommen zu werden.

Zur technischen Realisierung des Verzeichnisses werden verschiedene Softwaresysteme zum Betrieb von digitalen Bibliotheken eruiert. Eine bereits in der Antragsphase von re3data.org realisierte Testversion von re3data.org wird erweitert.

Die Entwicklung von re3data.org soll im Dialog mit etablierten Akteuren im Bereich des Forschungsdatenmanagements betrieben werden. Vor diesem Hintergrund wurde im Frühjahr 2012 ein Memorandum of Understanding mit DataCite<sup>24</sup> unterzeichnet. DataCite ist eine aus dem von der DFG-Projekt "Publikation und Zitierbarkeit wissenschaftlicher Primärdaten" (STD-DOI)<sup>25</sup> entstandene Initiative zur dauerhaften Adressierung von Forschungsdaten. Im Rahmen der Zusammenarbeit soll insbesondere der Informationsfluss zwischen den beiden Partnern gesichert werden.

---

<sup>22</sup> Siehe hierzu auch die Einführung von Klump (2011) sowie die im Rahmen des EU-Projektes APARSEN entstandenen Reports rund um die Qualitätssicherung von Forschungsdaten, z.B. Pampel et al. (2012).

<sup>23</sup> Zugang meint hier den Zugang zum Forschungsdaten-Repositorium, den darin vorgehaltenen Forschungsdaten sowie den Zugang zum Daten-Upload.

<sup>24</sup> <http://datacite.org>

<sup>25</sup> Eine Fortsetzung der Kooperation der damals beteiligten Partner wird seit August 2011 im Rahmen des DFG-Projektes „Kompetenzzentrum für Forschungsdaten aus Erde und Umwelt“ (KomFor) umgesetzt. Siehe dazu: <http://www.komfor.net>.

#### 4. Fazit

„We start the era of open science.“ Dieses Zitat der EU-Kommissarin für die Digitale Agenda Neelie Kroes (2012a) macht deutlich, dass Offenheit zum Paradigma der digitalen Wissenschaft wird. Zur Förderung dieser Entwicklung wird eine dauerhafte Informationsinfrastruktur benötigt, die Wissenschaftler beim „sharing“ ihrer Daten unterstützt und auch kommenden Forschergenerationen Zugang und Nachnutzung der Daten garantiert. Mitte 2012 befindet sich re3data.org nach einer sechsmonatigen Konzeptionsphase in der Realisierungsphase. Im Fokus steht dabei die technische Umsetzung des Verzeichnisses. Über den Aufbau des Verzeichnisses hinaus, widmet sich das Projekt der Vernetzung und Standardisierung von Forschungsdaten-Repositories und leistet so einen Beitrag auf dem herausfordernden Weg zu einer offenen Wissenschaft.

#### Literaturverzeichnis

Alle Links wurden am 09.07.2012 geprüft.

Allianz der deutschen Wissenschaftsorganisationen. (2010, June). Grundsätze zum Umgang mit Forschungsdaten. Retrieved from <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze/>

Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 40(Database issue), D48-D53. doi:10.1093/nar/gkr1202

Berliner Erklärung. (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Retrieved from <http://oa.mpg.de/lang/de/berlin-prozess/berliner-erklarung/>

Braun, K., Buddenbohm, S., Dobratz, S., Herb, U., Müller, U., Pampel, H., Schmidt, B., et al. (2010). *DINI-Zertifikat Dokumenten- und Publikationsservice 2010* (3.0 ed.). Göttingen. doi:urn:nbn:de:kobv:11-100109986

Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12(Suppl 15), S2. doi:10.1186/1471-2105-12-S15-S2

Consultative Committee for Space Data Systems (CCSDS). (2011). *Audit and Certification of Trustworthy Digital Repositories. Recommended Practice. CCSDS 652.0-M-1. Magenta Book. Practice*. Retrieved from <http://public.ccsds.org/publications/archive/652x0m1.pdf>

Cravedi, K. (2008). *GenBank Celebrates 25 Years of Service with Two-Day Conference. Leading Scientists Will Discuss the DNA Database at April 7-8 Meeting*. Retrieved from <http://www.nih.gov/news/health/apr2008/nlm-03.htm>

- Dallmeier-Tiessen, S, Dobratz, S., Gradmann, S., Horstmann, W., Kleiner, E., Pampel, H., Schirmbacher, P., et al. (2009). *Positionspapier Forschungsdaten* (1.0 ed.). doi:urn:nbn:de:kobv:11-10098082
- Dallmeier-Tiessen, Sünje. (2011). *Strategien bei der Veröffentlichung von Forschungsdaten*. Berlin. Retrieved from [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_173.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_173.pdf)
- Data Seal of Approval. (2010). *Data Seal of Approval. Guidelines version 1*. Retrieved from <http://assessment.datasealofapproval.org/documentation/>
- Deutsche Forschungsgemeinschaft. (2012). *Leitfaden für die Antragstellung. DFG-Vordruck 54.01 - 1/12*. Retrieved from [http://www.dfg.de/formulare/54\\_01/54\\_01\\_de.pdf](http://www.dfg.de/formulare/54_01/54_01_de.pdf)
- Diepenbroek, M., Grobe, H., Reinke, M., Schlitzer, R., & Sieger, R. (1999). Data management of proxy parameters with PANGAEA. In G. Fischer & G. Wefer (Eds.), *Use of Proxies in Paleocceanography. Examples from the South Atlantic* (pp. 715–727). Berlin, Heidelberg & New York: Springer-Verlag. Retrieved from <http://hdl.handle.net/10013/epic.11224>
- ELIXIR. (n.d.). *The ELIXIR Strategy for Data Resources. Draft Report from Workpackage 2. The ELIXIR Preparatory Phase*. Retrieved from [http://www.elixir-europe.org/bcms/elixir/Documents/reports/ELIXIR Strategy for Data Resources report.pdf](http://www.elixir-europe.org/bcms/elixir/Documents/reports/ELIXIR%20Strategy%20for%20Data%20Resources%20report.pdf)
- ESF & EUROHORCS. (2011). *Basic Requirements for Research Infrastructures in Europe*. Retrieved from [http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic\\_requirements\\_research\\_infrastructures.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requirements_research_infrastructures.pdf)
- Edinburgh DataShare. (n.d.). *What is Edinburgh DataShare?* (pp. 6-8). Retrieved from <http://datashare.is.ed.ac.uk>
- European Commission. (2009). ICT infrastructures for e-science. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2009) 108 final. Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF>
- European Strategy Forum on Research Infrastructures. (2006). *European Roadmap for Research Infrastructures. Report 2006*. Retrieved from [ec.europa.eu/research/infrastructures/pdf/esfri/esfri\\_roadmap/roadmap\\_2006/esfri\\_roadmap\\_2006\\_en.pdf](http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf)
- Galperin, M. Y., & Fernández-Suárez, X. M. (2012). The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Research*, 40(Database issue), D1-D8. doi:10.1093/nar/gkr1196
- High Level Expert Group on Scientific Data. (2010). *Riding the wave. How Europe can gain from the rising tide of scientific data*. Retrieved from <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- Hrynaskiewicz, I. (2012, April). LabArchives and BioMed Central: a new platform for publishing scientific data. *BioMed Central Blog*. Retrieved from [http://blogs.openaccesscentral.com/blogs/bmcblog/entry/labarchives\\_and\\_biomed\\_central\\_a](http://blogs.openaccesscentral.com/blogs/bmcblog/entry/labarchives_and_biomed_central_a)
- ICSU World Data System (WDS). (2011). *Certification of World Data System Members Introduction*.
- Kindling, M. (2009). *Möglichkeiten der Strukturmodellierung: Eine exemplarische Zusammenführung funktionaler Anforderungen an die Bereitstellung digitaler Forschungsdaten für ausgewählte geisteswissenschaftliche Disziplinen*. Humboldt-Universität zu Berlin. Retrieved from <http://www.nbn-resolving.org/urn:nbn:de:kobv:11-100185124>



- Klump, J. (2011). Criteria for the Trustworthiness of Data Centres. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-klump
- Klump, J., & Conze, R. (2007). The Scientific Drilling Database (SDDb) - Data from Deep Earth Monitoring and Sounding. *Scientific Drilling*, (4), 30-31. doi:10.2204/iodp.sd.4.06.2007
- Kommission Zukunft der Informationsinfrastruktur. (2011). *Gesamtkonzept für die Informationsinfrastruktur in Deutschland*. Retrieved from [http://www.allianz-initiative.de/fileadmin/user\\_upload/KII\\_Gesamtkonzept.pdf](http://www.allianz-initiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf)
- Kroes, N. (2012a). *Opening Science Through e-Infrastructures*. SPEECH/12/258. Retrieved from [europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258](http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258)
- Kroes, N. (2012b). *Making Open Access a reality for Science*. SPEECH/12/392. Retrieved from <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392&format=HTML&aged=0&language=EN&guiLanguage=en>
- Kuipers, T., & Van der Hoeven, J. (2009). *Insight into digital preservation of research output in Europe. Survey Report*. Retrieved from [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf)
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and Peer Review of Data: Moving Towards Formal Data Publication. *International Journal of Digital Curation*, 6(2). doi:10.2218/ijdc.v6i2.205
- Marcial, L. H., & Hemminger, B. M. (2010). Scientific Data Repositories on the Web : An Initial Survey. *Journal of the American Society for Information Science*, 61(10), 2029-2048. doi:10.1002/asi
- National Science Foundation. (2011a). *Award and Administration Guide. Chapter VI - Other Post Award Requirements and Considerations*. Retrieved from [http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag\\_6.jsp#VID4](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4)
- National Science Foundation. (2011b). *Grant Proposal Guide. Chapter II - Proposal Preparation Instructions*. Retrieved from [http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg\\_2.jsp#dmp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp)
- Nature. (2011). *Guide to Publication Policies of the Nature Journals*. Retrieved from <http://www.nature.com/authors/gta.pdf>
- Organisation for Economic Co-operation and Development (OECD). (2007). *Principles and Guidelines for Access to Research Data from Public Funding*. Paris. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- PANGAEA. (n.d.). About / Imprint. Retrieved July 29, 2011, from <http://www.pangaea.de/about/>
- Pampel, H., & Bertelmann, R. (2011). „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung. In S. Büttner, H.-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 49-61). Bad Honnef: Bock + Herchen. doi:urn:nbn:de:kobv:525-opus-2287
- Pampel, H., Pfeiffenberger, H., Schäfer, A., Smit, E., Pröll, S., & Bruch, C. (2012). *Report on Peer Review of Research Data in Scholarly Communication*. Retrieved from <http://epic.awi.de/30353/>
- Rice, R. (2009). *DISC-UK DataShare*. Retrieved from <http://ie-repository.jisc.ac.uk/336/>
- Schaaf, I. (2011). *Forschungsdaten im Netz – Untersuchung von online verfügbaren Repositorien*. Hochschule der Medien Stuttgart.

- Schallehn, V. (2010, October 13). Open Data LMU. *Universitätsbibliothek München*. Retrieved from [http://www.ub.uni-muenchen.de/no\\_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/](http://www.ub.uni-muenchen.de/no_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/)
- Schindler, U., Diepenbroek, M., & Grobe, H. (2012). PANGAEA - Research Data enters Scholarly Communication. *Geophysical Research Abstracts*, 14, EGU2012-13378-1. Retrieved from <http://meetingorganizer.copernicus.org/EGU2012/EGU2012-13378-1.pdf>
- Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Kindling, M., Pampel, H., Manova, S., et al. (2012). *Vocabulary for the Registration and Description of Research Data Repositories. Version 1.0*. doi:10.2312/re3.001
- Woutersen-Windhower, S., & Brandsma, R. (2009). Enhanced Publications, State of the Art. In M. Vernooy-Gerritsen (Ed.), *Enhanced Publications. Linking Publications and Research Data in Digital Repositories*. Amsterdam: Amsterdam University Press.
- Van der Graaf, M., & Waaijers, L. (2011). *A Surfboard for Riding the Wave. Towards a four country action programme on research data*. Retrieved from [http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=/Files/Filer/downloads/Primary+Research+Data/Surfboard+for+Riding+the+Wave/KE\\_Surfboard\\_Riding\\_the\\_Wave\\_Screen.pdf](http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=/Files/Filer/downloads/Primary+Research+Data/Surfboard+for+Riding+the+Wave/KE_Surfboard_Riding_the_Wave_Screen.pdf)

---

## **TEODOOR: Geodateninfrastruktur zur Verwaltung und Veröffentlichung von terrestrischen Beobachtungsdaten der HGF Infrastrukturmaßnahme TERENO aus verteilten Quellen**

Ralf Kunkel<sup>1</sup>, Jürgen Sorg<sup>1</sup>, Rainer Gasche<sup>2</sup>, Jens Klump<sup>3</sup>, Olaf Kolditz<sup>4</sup>,  
Mark Frenzel<sup>4</sup>, Frank Neidl<sup>2</sup>

<sup>1</sup> Forschungszentrum Jülich GmbH (IBG-3)

<sup>2</sup> Karlsruher Institut für Technologie (KIT)

<sup>3</sup> Helmholtz-Zentrum Potsdam (GFZ)

<sup>4</sup> Helmholtz-Zentrum für Umweltforschung (UFZ)

### **Zusammenfassung**

Im Rahmen der HGF-Infrastrukturmaßnahme TERENO werden in Deutschland seit 2008 vier langfristig betriebene terrestrische Umweltobservatorien eingerichtet, in denen große Mengen an umweltrelevanten Daten gesammelt werden. Um diese Datenmengen adäquat zu verwalten, zu beschreiben, auszutauschen und zu veröffentlichen wurde die verteilte Geodateninfrastruktur TEODOOR (<http://www.tereno.net>) aufgebaut. Die für die Einrichtung und Betrieb der Observatorien verantwortlichen Institutionen betreiben dabei jeweils eigene Dateninfrastrukturen, die über lokale, standardisierte OGC-konforme Web-Dienste miteinander kommunizieren. Ein zentrales TEODOOR-Datenportal hat die Aufgabe, die Daten und Metadaten aus den einzelnen Observatorien zusammenzuführen und für die interne und öffentliche Verwendung bereitstellen. Eine Reihe von Werkzeugen wie z.B. die hierarchische Suche nach Stichworten, Web-GIS Funktionalitäten und Visualisierungswerkzeuge erlauben es dem Benutzer, Daten von verschiedenen Zeitreihen und Variablen zu suchen, zu selektieren, darzustellen und konform zur gemeinsamen Datenpolitik auch herunterladen werden. Derzeit sind über TEODOOR zeitaktuelle Daten von mehr als 450 automatischen Stationen frei verfügbar.

### **Abstract**

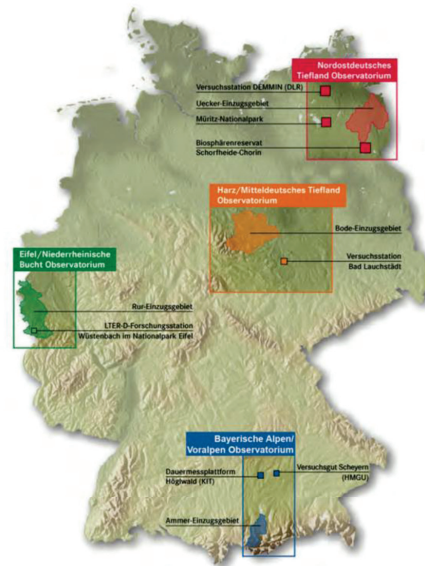
Within the TERENO initiative four long-term terrestrial observatories, which collect huge amounts of environmental relevant data, have been set up since 2008. In order to manage, describe, exchange and publish these data, the distributed Spatial Data Infrastructure TEODOOR (<http://www.tereno.net>) has been created. Each institution responsible for an individual observatory set up its own local data infrastructure, which may communicate to each other to exchange data and metadata internally or to the public by OGC-conformal Web services. The TEODOOR data portal serves as a database node to provide scientists and decision makers with reliable and well accessible data and data products. Various tools like hierarchical search or Web-GIS functions allow a deeper insight into the different observatories, test areas and sensor networks. Sensor data can be selected according to the measured parameters, stations and/or time periods, visualized and downloaded according to the common TERENO data policy. Currently, TEODOOR provides free access to data from more than 450 monitoring stations.

## Einleitung

Die Änderungen von Klima und Landnutzung sind wesentliche Faktoren des globalen Wandels, die nicht nur die Gesellschaft, sondern auch lebenswichtige Ökosystemfunktionen und Ressourcen stark beeinflussen. Langfristige Trends in Temperatur, Niederschlag und weiteren klimatischen Parametern beeinflussen alle Umweltkompartimente nicht nur direkt, sondern auch indirekt durch komplexe Rückkopplungsmechanismen. Eine wichtige Aufgabe der terrestrischen Ökosystemforschung ist die Entwicklung und Bereitstellung integrierter Modelle zur Quantifizierung der Auswirkungen von Umweltveränderungen auf terrestrische Systeme. Als Grundlage hierfür sind langfristige, interdisziplinäre und skalenübergreifende Umweltdaten notwendig (Bronstert et al., 2009; Reid et al., 2009; Richter and Mobley, 2009). In diesem Zusammenhang wird gerade in letzter Zeit die Entwicklung und Implementierung von Dateninfrastrukturen für die Umweltforschung sehr intensiv betrieben (Band et al., 2005; Burt et al., 2008; Keller et al., 2008; Lin, 2003; Lin, 2010; McDonnell et al., 2007; Montgomery et al., 2007; Nisbet, 2007; Parr et al., 2003; Zoback, 2001). Getrieben durch die rasante Entwicklung hochauflösender Mess- und Fernerkundungstechnik werden in diesen Infrastrukturen große Mengen heterogener Umweltdaten generiert.

Um diese Datenmengen für terrestrische Systemanalysen nutzbar zu machen, sind zwei Schritte erforderlich. Zum einen müssen neuartige Datenmanagement- und Schnittstellenkonzepte entwickelt werden, um die anfallenden Daten in den verschiedenen Zeit- und Raumskalen verarbeiten und effizient austauschen zu können. Hierzu müssen die existierenden, aktuell meist nicht kompatiblen Ansätze zur Schaffung von Interoperabilität in den einzelnen Wasser- und Klimaforschungs-Communities geeignet zusammengeführt, visualisiert und über standardisierte Webdienste publiziert werden. In diesem Zusammenhang spielen die Aktivitäten und Standards des *Open Geospatial Consortiums* (OGC) eine wichtige Rolle. Zum anderen müssen integrierte Modellansätze direkt an diese Datenströme und Informationsinfrastrukturen angebunden werden. Nur so kann es gelingen, das Systemverständnis unter Berücksichtigung der Wechselwirkungen von Atmosphäre, Landoberfläche, Gewässer, Boden bis hin zum Grundwasser auf einer so derzeit nicht verfügbaren Umweltdatenbasis auf ein neues Niveau zu heben.

Mit dem TERENO-Vorhaben (TERrestrial ENvironmental Observatories) hat die Helmholtz-Gemeinschaft 2008 begonnen, ein Netzwerk langfristig betriebener, integrierter Umweltobservatorien in Deutschland einzurichten (Bogena et al., 2006; Zacharias et al., 2011) mit dem Ziel, die Auswirkungen des Klima- und Landnutzungswandels auf die terrestrischen Umweltsysteme langfristig, interdisziplinär und skalenübergreifend zu untersuchen. Hierzu wurden in Deutschland vier terrestrische Observatorien eingerichtet, die von jeweils einem Helmholtz-Zentrum betrieben werden (siehe Abbildung 1).



**Abbildung 1:** Lage der vier TERENO-Observatorien (Zacharias et al., 2011)

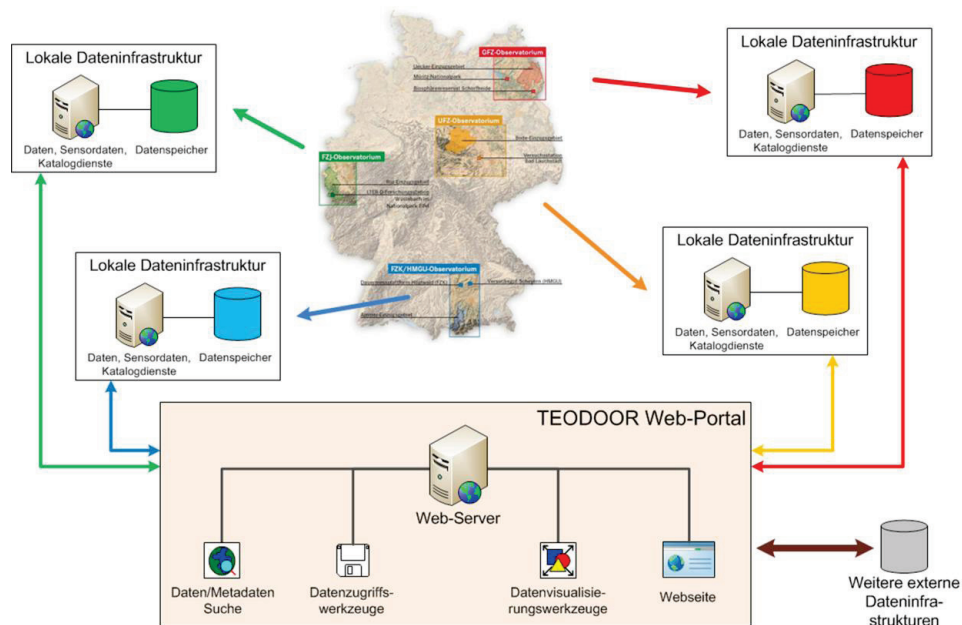
Übergeordnetes Ziel von TERENO ist letztlich die Nutzung der in den einzelnen Observatorien erhobenen Umweltdaten für prognosefähige Systemanalysen, um beispielsweise die Einflüsse etwaiger Klima- und Landnutzungsänderungen auf die langfristige Verfügbarkeit und Qualität von Wasserressourcen abschätzen zu können.

### Aufbau von TEODOOR

Der gesamte Erfolg von Langzeitprojekten wie TERENO hängt ganz entscheidend von einer gut organisierten Datenverwaltung, dem Datenaustausch zwischen den beteiligten Teilprojekten und der Verfügbarmachung der erhobenen Daten ab. Um dieses zu gewährleisten, wurde die räumliche Dateninfrastruktur (SDI) TEODOOR (TEReno Online Data RepOsitORY) aufgebaut, welche das Bindeglied bildet, über das Wissenschaftlern und Entscheidungsträgern zuverlässige und zugängliche Umweltdaten und weiterführende Produkte zur Verfügung gestellt werden.

Aufgrund der Heterogenität der Daten, der großen Menge anfallender Daten und der bereits bestehenden Dateninfrastrukturen in den beteiligten Institutionen wurde in

TERENO ein dezentrales Datenmanagementkonzept umgesetzt (siehe Abbildung 2). Die Grundidee hierbei ist die Verknüpfung lokaler unabhängiger Infrastrukturen (Daten- bzw. Metadatenbanken) in den einzelnen Observatorien mit einer zentralen TERENO-Portalanwendung (TEODOOR). Das TEODOOR Geodatenportal bündelt die über die verschiedenen Webdienste distributiv zur Verfügung gestellten Daten der einzelnen Observatorien über Such-, Visualisierungs- und Downloadwerkzeuge und stellt sie der wissenschaftlichen Gemeinschaft zur Verfügung. Hierzu werden Webdienste verwendet, die durch das Open Geospatial Consortium (OGC) standardisiert wurden (OGC, 2012), wie z.B. Web Coverage Services (WCS), Web Feature Services (WFS), Web Map Services (WMS), Web Processing Services (WPS), Web Catalogue Services (CSW) und insbesondere dem Sensor Web Enablement (SWE). Diese Webdienste stellen Standards und Schnittstellen zum Suchen in und zum Zugriff auf Geodatenbestände zur Verfügung. Durch die Verwendung standardisierter Schnittstellen wird zum einen eine größtmögliche Flexibilität des Datenmanagements in den einzelnen Observatorien sichergestellt. Zum anderen wird auch die Übertragbarkeit der Daten aus den einzelnen Observatorien gewährleistet, so dass auch Daten von weiteren, neu eingerichteten Observatorien oder anderen behördliche Datenhaltern ohne größeren Aufwand eingebunden werden können.



**Abbildung 2:** Aufbau der Dateninfrastruktur TEODOOR

Ein wesentlicher Aspekt von TERENO ist die Verfügbarmachung der gemessenen Daten für die Öffentlichkeit. Als Grundlage hierfür wurde in einem partizipatorischen Prozess eine gemeinsame TERENO-Datenpolitik erarbeitet, in der festgelegt wurde, welche Daten unter welchen Bedingungen zugänglich sind (TERENO, 2010). Innerhalb des TERENO-Verbundes sind alle Daten frei verfügbar. Der Öffentlichkeit werden grundsätzlich alle Daten umgehend zugänglich gemacht, wenn die Daten mindestens eine erste Qualitätskontrolle erfolgreich durchlaufen haben und der Veröffentlichung keine anderen Nutzungsvorbehalte, z.B. für laufende Promotionsvorhaben entgegenstehen. Neben der Datenqualität spielt insbesondere die adäquate Beschreibung der Datenbestände über Metadaten eine entscheidende Rolle. Aus diesem Grunde werden Daten immer als eine Kombination der Daten und der beschreibenden Metadaten behandelt. Durch ein gemeinsames TERENO-Metadatenprofil wurde eine einheitliche Datenbeschreibung innerhalb des TERENO-Verbundes sichergestellt.

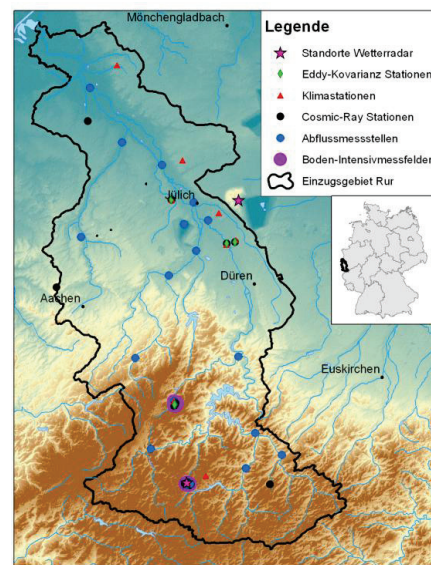
Während der Einrichtung und Instrumentierung der TERENO-Observatorien wurden von den beteiligten Zentren Dateninfrastrukturen eingerichtet, in welche die TERENO-Daten sukzessive integriert werden. Auch wenn die vier Observatorien



Unterschiede hinsichtlich des Schwerpunktes der untersuchten Fragestellungen aufweisen, sind die anfallenden Daten bei den verschiedenen Observatorien sehr ähnlich und sollen am Beispiel der Zeitreihendaten des vom FZJ betriebenen Eifel-Rur Observatoriums exemplarisch verdeutlicht werden.

#### *Lokale Dateninfrastruktur „Eifel-Rur Observatorium“*

Das Observatorium Eifel-Rur umfasst das Einzugsgebiet der Rur mit einer Größe von 2354 km<sup>2</sup>. In drei repräsentativen Regionen des Einzugsgebietes wurden Intensivmessgebiete eingerichtet, in denen eine Reihe von Parametern, wie z.B. Bodenfeuchte und Bodentemperaturen, Grundwasserständen sowie Energie und Gasflüssen erfasst werden. Verteilt über das Einzugsgebiet werden zusätzlich an Cosmic-Ray Stationen Bodenfeuchtedaten (Desilets et al., 2010) sowie an Eddy-Kovarianz Stationen Klima-, Energie- und Gasflussdaten erhoben (siehe auch Abbildung 3):



**Abbildung 3:** Lage der Intensivtestgebiete im Eifel-Rur Observatorium.

- 53 automatische Stationen liefern mit einer Frequenz von 1-5 h<sup>-1</sup> Abfluss-, Klima-, Oberflächengewässerqualitäts- und Bodendaten. Pro Tag werden von diesen Stationen insgesamt 26400 einzelne Datenwerte erfasst.
- Zwei SoilNet Beobachtungsnetzwerke (Bogena et al., 2010) erfassen mit einer Frequenz von 6 h<sup>-1</sup> Bodentemperaturen, Bodenfeuchten und Klimaparameter an insgesamt 404 Messstellen mit einem Umfang von insgesamt 686000 Messwerten pro Tag.
- Durch sieben Eddy-Kovarianz Stationen (Vanderborgh et al., 2010) werden eine Reihe von Klimaparametern und Gaskonzentrationen mit einer Frequenz von 20 s<sup>-1</sup> (Rohdaten) bzw. 6 h<sup>-1</sup> (prozessierte Daten) gemessen, die sich pro Tag auf 133 Millionen einzelner Messwerte summieren.

- Ein Wetterradar auf der Sophienhöhe nahe Jülich und ein kleinerer Rainscanner im Süden des Einzugsgebietes erzeugen mit einer Frequenz von  $12\text{ h}^{-1}$  bzw.  $60\text{ h}^{-1}$  jeweils ein Niederschlagsraster mit  $800 \times 800$  Pixeln. Pro Tag fallen 1728 Rasterdateien an, die zusammen mit den Rohdaten insgesamt 4.3 GB Speicherplatz beanspruchen.

Zusätzlich werden Abfluss- und Klimadaten aus externen Beobachtungsmessnetzen von Wasserverbänden und dem Bundesland NRW integriert. Neben den Stationsdaten fallen auch Dokumente, Bilder und GIS-Daten in größerem Umfang an. Diese Daten werden über ein Content-Management System (Plone) und einen GIS-Server verarbeitet, was allerdings hier nicht weiter betrachtet werden soll.

#### Stationsdaten

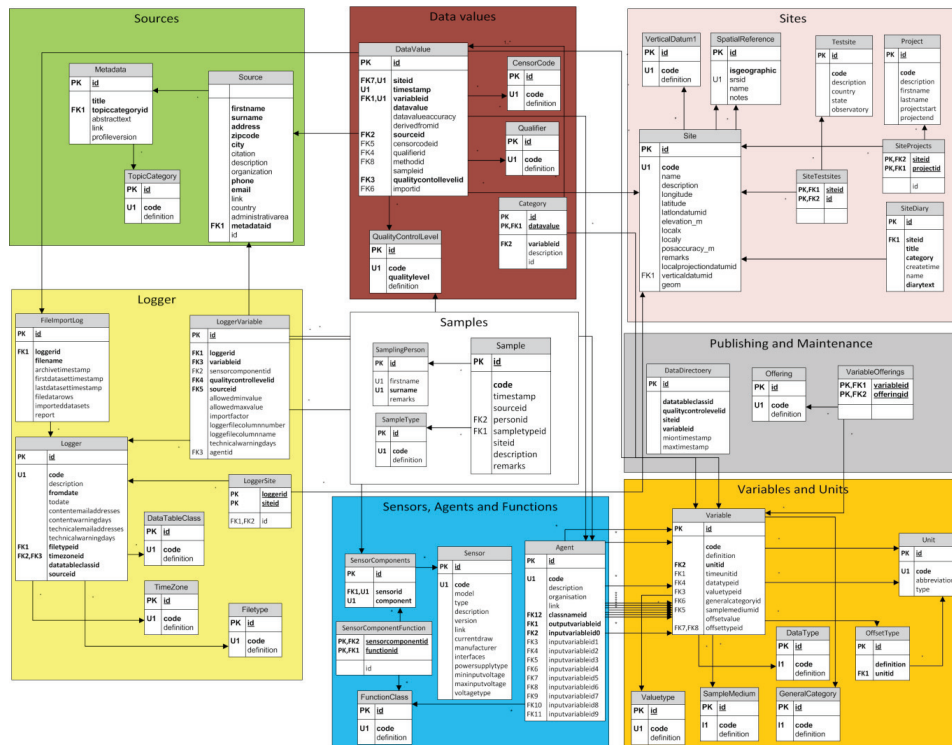
Die Verarbeitung der Stationsdaten erfolgt automatisiert durch ein für diesen Zweck im FZJ entwickeltes integriertes Zeitreihenmanagementsystem (TSM 2.0). Die Sensordaten, welche die Infrastruktur über unterschiedliche Kanäle, wie z.B. DSL, E-Mail, GSM, Richtfunk oder manuellen Upload erreichen, werden in einem ersten Schritt in ASCII-kodierte Dateien überführt. Der Inhalt der Dateien wird in einem nächsten Schritt in eine relationale Datenbank überführt, wobei der Überführungsprozess für die einzelnen Stationen bzw. Stationstypen individuell konfigurierbar ist. Gegebenenfalls erfolgt an dieser Stelle auch eine Umrechnung der Rohdaten, z.B. von Signalstärken in Temperaturen, die Anwendung von Umrechnungsvorschriften für mehrere Eingangsgrößen, wie z.B. die Bildung von Mittelwerten oder zeitlichen Aggregaten. Konfigurierbare Filter ermöglichen es, eine erste Plausibilitätsprüfung und Kennzeichnung der Datenqualität vorzunehmen. Nach erfolgreichem Datenimport werden die Originaldateien archiviert. Das System benachrichtigt die für eine Messstelle verantwortlichen Techniker und/oder Wissenschaftler über den erfolgten Importprozess, über aufgetretene Fehler, Warnungen und Datenlücken, aber auch über Probleme bei der Übertragung der Rohdaten von der Messstelle zur Infrastruktur.

Gemessene Daten sind primär durch Ort und Zeitpunkt der Messung, die physikalische Größe und den Messwert charakterisiert. Für eine spätere Nutzung der Daten sind diese Angaben allein allerdings nicht ausreichend. Hierzu sind zusätzliche beschreibende Angaben, die sogenannten Metadaten, erforderlich (Bose, 2002; Gray et al., 2005; Michener et al., 1997; Tomasic and Simon, 1997).

Metadaten enthalten alle notwendigen Angaben zur Charakterisierung der gemessenen Größe, wie z.B. Namen, Attribute, Einheiten, Variablenbeschreibungen, Datengenauigkeit, Datenlayout, beschreiben aber auch den Prozess der Messung, Aufbereitung und Berechnung der Daten (Gray et al., 2005). Um dem Rechnung tragen zu können, wurde ein Datenmodell für Zeitreihendaten entwickelt, in dem neben den eigentlichen Daten auch alle relevanten Metadaten verwaltet werden können. Das Datenmodell basiert in weiten Teilen auf dem Observation Data Model (ODM), das im Rahmen der CUAHSI-Initiative entwickelt und eingesetzt wird (Horsburgh et al., 2008; Tarboton et al., 2009). Es wurde hier allerdings erweitert, um die Prozesse von Datenerfassung, Datenimport und Datenveröffentlichung adäquat abbilden zu können.

Eine schematische Darstellung des Datenmodells gibt Abbildung 4. Es beinhaltet mehrere Komponenten, über die eine detaillierte Charakterisierung der Messwerte (data values), der Messstellen (sites) der Messgrößen (variables) erfolgt. Zusätzlich werden die bei der Messung verwendeten Sensoren, die Konfiguration der einzelnen Logger und die beim Import anzuwendenden Berechnungsverfahren beschrieben (logger, sensors, agents and functions). Das Modell enthält darüber hinaus Tabellen zur Spezifizierung der jeweils verantwortlichen Personen (sources), zur Verwaltung von Proben (samples) sowie zur Veröffentlichung der Daten (publishing and maintenance).

Neben der Langzeitspeicherung und Dokumentation der Sensordaten ist insbesondere die Schaffung von Zugängen zu den Daten ein wesentlicher Aspekt von TERENO. Innerhalb des distributiven Aufbaus von TERENO stellt dies nicht zuletzt aufgrund der großen Vielfalt an Herstellern von Sensorsystemen und den unterschiedlichen Übertragungsprotokollen eine große Herausforderung dar. Der Aufbau von Beobachtungsnetzwerken erfordert eine interoperable, hardware- und plattformunabhängige Behandlung der einzelnen Sensoren (Bröring et al., 2011a; Bröring et al., 2011b; Jirka et al., 2012), so dass die Daten kohärent beschrieben, gesucht, ggf. kontrolliert sowie über einheitliche Schnittstellen auf sie zugegriffen werden kann.



**Abbildung 4:** Datenmodell zur Verarbeitung der Sensordaten im TERENO-Observatorium Eifel-Rur.

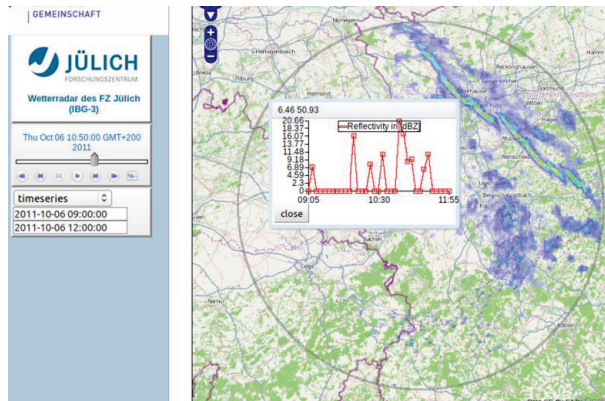
Eine solche einheitliche Behandlung von Sensoren kann durch den vom OGC bereitgestellten Sensor Web Enablement (SWE) realisiert werden, welcher eine Reihe aufeinander abgestimmter Standards und Schnittstellendefinitionen zur Verfügung stellt (Botts et al., 2008): der Observation & Measurement Standard zur Beschreibung und zum Austausch von Beobachtungsdaten (Cox, 2010; Cox, 2011), die Sensor Model Language (SensorML) als Informationsmodell zur Beschreibung von Sensoren und der von ihnen gelieferten Informationen (Botts, 2007) und die TransducerML (TML) als ein Protokoll für den Austausch von Informationen zwischen den Sensoren und einer SensorWeb-Infrastruktur. Als Dienste definiert der SWE den Sensor Observation Service (SOS) für die Suche, das Filtern und den Abruf von Sensordaten (Bröring et al., 2012; Na and Priest, 2007), den Sensor Planning Service (SPS) zum Abrufen und Web-basierten Steuern von Sensoren (Simonis and Echterhoff, 2011), den Sensor Alert Service (SAS), der es Nutzern ermöglicht, über eine Standardschnittstelle Alarmnachrichten von einzelnen Sensoren zu erhalten und

schließlich den Web Notification Service (WNS) (Simonis and Wytzisk, 2003) für den asynchronen Austausch von Informationen zwischen den einzelnen Webdiensten.

Die Open-Source Initiative 52North hat die SWE-Dienste implementiert und als freie Software veröffentlicht (52°North, 2012). Im Rahmen von TERENO werden die Sensordaten über die 52North Implementierung zugänglich gemacht, wobei zum einen der interne Zugriff auf die Daten an das verwendete Datenmodell angepasst und zum anderen in das O&M-Modell zum Datenaustausch TERENO-relevanten Angaben, wie z.B. die Qualitätskennzeichnungen, aufgenommen wurden. Für das Observatorium Eifel-Rur wurde für jedes Intensivmessgebiet und das gesamte Einzugsgebiet jeweils ein Dienst für interne und externe Nutzung aufgesetzt, so dass die Sensordaten durch insgesamt 14 Sensor Observation Services veröffentlicht werden.

#### Rasterzeitreihen

Die von einem Sensor gemessenen Werte werden einem Geoobjekt zugeordnet, dem sogenannten *Feature Of Interest* (FOI). Dieses Geoobjekt, für das die Messwerte gelten, kann der Punkt der Messung sein, aber auch aus einer Fläche oder einem Volumen bestehen. Die bisher betrachteten Messdaten hatten alle zu Eigen, dass die mit einem Sensor zu einem bestimmten Zeitpunkt gemessenen Größen ein Integral für das gesamte Feature-Of-Interest repräsentieren. Sensoren können aber auch für einen Zeitpunkt flächendifferenzierte Daten, meist Rasterdaten liefern. Hierzu zählen alle Arten von Fernerkundungsdaten, bei denen bei einem Überflug eines Flugzeugs oder Satelliten die flächenhafte Variation einer Zustandsgröße (Landnutzung, Radarecho etc.) für ein bestimmtes Gebiet gemessen wird. Hierzu zählen aber auch die Daten, die von den im TERENO-Vorhaben eingesetzten Radargeräten erzeugt werden. Diese liefern in regelmäßigen Zeitabständen (1-5 Minuten) jeweils einen Rasterdatensatz mit Reflektivitäten, aus denen über Berechnungsvorschriften u.a. Niederschlagshöhen ermittelt werden können. Auch diese Rasterzeitreihendaten sind in der OGC-Spezifikation von Sensor Observation Services vorgesehen; es existierte bislang aber keine Implementierung des Standards. Aus diesem Grunde wurde auf Grundlage des SOS des 52North ein OGC konformer Sensor Observation Service entwickelt, der auch für Rasterzeitreihendaten anwendbar ist (Sorg, 2012). Hierzu wurde ein Datenmodell entwickelt, welches dem hohen Speicherbedarf der Rasterdaten (hier etwa 65000



**Abbildung 5:** Klientanwendung zur animierten Darstellung von Rasterzeitreihen am Beispiel des Wetterradars Sophienhöhe.

mal größer als Punktdaten) Rechnung trägt, es aber auch ermöglicht, die Daten in Bilder mit variabler Farbskala umzuwandeln. Darüber hinaus wurden effiziente Algorithmen zur räumlichen Suche und zur Filterung von Rasterdaten entwickelt. Damit wurde im Ergebnis ein OGC-konformer Dienst zu Verfügung gestellt, der einen

Zugang zu einzelnen Reflektivitätsrastern und frei wählbaren Ausschnitten bis hin zu Zeitreihendaten an frei wählbaren Einzelpunkten ermöglicht. Durch die Kopplung mit einem Sensor Alarm Service (SAS) wurden die Möglichkeiten eines SOS zur Filterung von Daten genutzt, in dem Nutzern die Möglichkeit geschaffen wurde, sich sektorbezogen vor Starkregenereignissen warnen zu lassen (Sorg, 2012).

Zur Visualisierung von Rasterzeitreihendaten wurde eine Klientanwendung entwickelt, die ausschließlich über die OGC-Standardschnittstellen mit einem Raster-SOS kommunizieren kann. Die Anwendung beinhaltet einen Player zur Darstellung animierter Rasterzeitreihen aus einem wählbaren Zeitraum mit wählbarer räumlicher Ausdehnung und die Möglichkeit, sich eine Zeitreihe der Rasterdaten an einem wählbaren Ort ausgeben zu lassen. Dies ist in Abbildung 5 am Beispiel einer Zeitreihe von Radarreflektivitäten des Wetterradars Sophienhöhe dargestellt. Der Klient ist in das TEODOOR-Datenportal eingebunden und wird dort derzeit zur Visualisierung von Wetterradardaten eingesetzt<sup>1</sup>, kann aber für alle Arten von Rasterzeitreihendaten, die über einen Raster-SOS zur Verfügung gestellt werden, verwendet werden.

<sup>1</sup> [http://teodoor.icg.kfa-juelich.de/@@ibg.coverage\\_player](http://teodoor.icg.kfa-juelich.de/@@ibg.coverage_player)

### Datensuche und Katalogdienst

Eine Reihe von Untersuchungen haben gezeigt, dass die Nachfrage an distributiv verfügbaren Daten zwar kontinuierlich ansteigt, ihre tatsächliche Nutzbarkeit aber durch mangelnde Kenntnis der vorhandenen Daten, durch schlechte Dokumentation der Daten und insbesondere durch Dateninkonsistenzen eingeschränkt wird (Nogueras-Iso et al., 2005; Tomasic and Simon, 1997). Eine weitergehende Nutzung der in einer Dateninfrastruktur verwalteten Daten und der zur Verfügung gestellten Datendienste erfordert zwingend deren umfassende Beschreibung durch Metadaten. In Abhängigkeit vom Datentyp existieren gängige Standards, die festlegen, durch welche Angaben die Daten und die Datendienste beschrieben werden (FGDC, 2000; ISO, 2003; ISO, 2009). Für einzelne Standards werden oftmals ergänzend Profile definiert, mit denen die zwingend oder optional geforderten Elemente näher spezifiziert, festgelegt oder eingeschränkt werden. Auch im Rahmen des TERENO-Vorhabens wurde ein Metadatenprofil definiert, welches auf der EU-INSPIRE-Richtlinie Metadaten (ECJRC, 2009) basiert.

Katalogdienste ermöglichen es Nutzern oder anderen Geodiensten, herauszufinden, welche Datendienste verfügbar sind und welche für die Nutzer relevanten Daten sie enthalten. In der Katalogdienst-Spezifikation des OGC (Nebert et al., 2007) ist festgelegt, wie diese Katalogdienste aufgebaut sind und auf welche Art und Weise Metadaten gesucht und übertragen werden. Katalogdienste werden für die meisten gängigen Typen von Geodaten, aber auch für nicht raumbezogenen Daten standardmäßig eingesetzt.

Bei der Kopplung von Katalogdiensten mit Sensor(meta)daten gibt es derzeit allerdings noch eine Reihe von Problemen, die aus der besonderen Charakteristik von Sensornetzwerken resultieren. An erster Stelle spielt hierbei die Tatsache eine Rolle, dass bei Sensornetzwerken Zeitreihen von meist mehreren Parametern anfallen. Aufgrund der Dynamik der Datenerfassung ist dies durch die gängigen Metadatenstandards kaum abbildbar. Weiterhin werden die Sensormetadaten, d.h. Informationen zu den einzelnen Sensoren, den Messprozessen, den gemessenen Eingabe- und Ausgabegrößen sowie andere wichtige Informationen entsprechend dem SWE durch die SensorML spezifiziert. Diese ist allerdings nur eingeschränkt kompatibel zu den für allgemeine Geodaten verwendeten Standards (ISO, FGDC, DublinCore). Schließlich bestehen bei Sensornetzwerken besondere Anforderungen



an die Suche nach Informationen, z.B. nach bestimmten Messgrößen, Messorten, Zeitpunkten der Messung, oder Datengruppen (Jirka et al., 2009). Diese Informationen sind ebenfalls in der SensorML spezifiziert und über Anfragen an einen Sensor Observation Services zugänglich, aufgrund der unterschiedlichen Strukturen nicht jedoch über die klassischen Protokolle der Web-Katalogdienste.

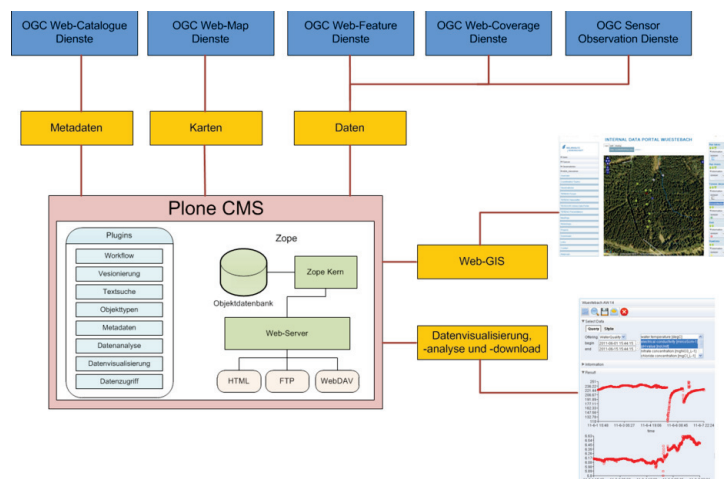
Eine Lösung für dieses Problem bietet das Electronic Business Registry Information Model, kurz ebRIM, welches ein generelles XML-basiertes Modell zur Verwaltung elektronischer Daten darstellt (OASIS, 2002). ebRIM stellt dabei nur eine Grundstruktur zur Verfügung, die mittels einer Erweiterung für SensorML-Metadatendokumente (Chen et al., 2009) anwendbar gemacht wurde. Um die sehr umfangreichen Möglichkeiten, Sensoren mittels SensorML zu beschreiben, auf ein Minimum an obligatorischen Werten zu beschränken, wird ein Profil der SensorML definiert, dass aus Elementen zur Identifikation und Klassifikation sowie aus räumlichen und zeitlichen Elementen besteht. Dadurch wird sichergestellt, dass diese obligatorischen Elemente in jedem auf ebRIM basierenden Katalogdienst (CSW) verfügbar sind, der das SensorML-Profil unterstützt. Realisiert wurde der auf ebRIM basierende Katalogdienst durch das quelloffene Buddata ebXML Registry/Repository (Buddata-ebRR) Projekt, dass den OASIS ebXML Registry und den OGC CSW Standard implementiert (kZen Labs, 2012). Aus den vorhandenen SensorML-Dokumenten wurden mittels xslt-Transformationen ebRIM-XML-Dokumente erzeugt und über die vom Buddata-ebRR-Dienst bereitgestellten Schnittstellen in das ebRIM überführt. Im TEODOOR Portal wurde eine benutzerfreundliche, grafische Schnittstelle implementiert, die eine Suche nach ebRIM-Metadaten, z.B. nach Sensornamen, Parametern, Messzeiträumen oder Datengruppen im Buddata-ebRR-Dienst durch Nutzung der dort implementierten standardisierten CSW-Operationen (Lesage, 2007) ermöglicht.



### TEODOOR Datenportal

Zusätzlich zu seiner Funktion als gemeinsame Informations- und Austauschplattform für den TERENO-Verbund hat das TEODOOR-Portal die Funktion, die in den einzelnen Observatorien gemessenen und über OGC-Dienste zur Verfügung gestellten Daten zusammenzuführen und der Öffentlichkeit zugänglich zu machen. In Abbildung 6 ist der Aufbau des TERENO Datenportals TEODOOR dargestellt. Im Kern besteht das System aus dem Content Management System Plone (Plone Foundation, 2004), welches die TERENO Webseite zur Verfügung stellt und es gleichzeitig TERENO Nutzern erlaubt, eine große Vielfalt von Objekttypen (Dokumente, Daten, Termine, Ressourcen, ...) in das System zu laden, zu verwalten und zu teilen. Plone verfügt über umfangreiche Such- und Datenzugriffsfunktionalitäten und unterstützt nativ den DublinCore Metadatenstandard (ISO, 2009).

Der überwiegende Teil der zur Verfügung stehenden Daten, insbesondere die Geo und Sensordaten, werden jedoch nicht über TEODOOR verwaltet, sondern sind über OGC-Schnittstellen an entfernte Datenquellen angekoppelt. Hierzu wurde das System erweitert, so dass eine Datensuche in verteilten OGC-Katalogdiensten möglich ist. Diese umfasst die hierarchische Suche nach Schlüsselworten, die Suche in Textdokumenten, aber insbesondere auch die Suche nach Sensordaten in entfernten ebRIM-Metadatenbeständen. Der Datenzugriff erfolgt direkt über die in den Metadaten angegebenen Speicherorte (Online Resources).



**Abbildung 6:** Aufbau des TEODOOR Datenportals.

Zur Darstellung und Visualisierung der räumlichen Daten wurden in TEODOOR WebGIS Funktionalitäten integriert. Diese ermöglichen es, die über entfernte Kartendienste (WMS) zur Verfügung gestellten Daten (Raster- oder Vektordaten) grafisch darzustellen. Darüber hinaus wurden Anwendungen entwickelt, um die Daten der Sensornetzwerke (alle oder Suchergebnisse) von entfernten SOS abzurufen, darzustellen und zu visualisieren. Auch hierbei wurden ausschließlich die OGC-Standards verwendet. Der Nutzer kann hierzu in einer WEBGIS Anwendung die für ihn relevanten Messstellen selektieren, und über eine Klientanwendung (siehe Abbildung 7) Parameter und Zeitreihe auswählen. Die Daten werden dann als Grafiken dargestellt und können im O&M-Format (Cox, 2011) heruntergeladen werden. Entsprechend der TERENO Datenpolitik sind die Daten grundsätzlich frei verfügbar; sie werden den Nutzern nach Anerkennung der Nutzungsbedingungen per E-Mail zugestellt, wobei die für eine bestimmte Messstelle verantwortlichen Wissenschaftler parallel über den Datenabruf benachrichtigt werden.

### Zusammenfassung und Ausblick

Im Rahmen der HGF-Infrastrukturmaßnahme TERENO werden in vier terrestrischen Umweltobservatorien große Mengen an umweltrelevanten Daten gesammelt. Um diese Datenmengen adäquat zu verwalten, zu beschreiben, auszutauschen und zu veröffentlichen, wurde die verteilte Dateninfrastruktur TEODOOR eingerichtet. Die für die Einrichtung und Betrieb der Observatorien verantwortlichen Institutionen betreiben dabei jeweils eigene lokale Dateninfrastrukturen, in denen effektive Werkzeuge für eine automatisierte Datenspeicherung und Datenarchivierung sowie für regelmäßige Plausibilitätsprüfungen der Daten zum Einsatz kommen.



**Abbildung 7:** Visualisierung von Sensordatenzeitreihen in TEODOOR am Beispiel der Klima- und Cosmic-Ray Station Gevenich im Rureinzugsgebiet.

Zusätzlich zu seiner Funktion als gemeinsame Informations- und Austauschplattform für den TERENO-Verbund hat das TEODOOR-Portal (<http://www.tereno.net>) die Funktion, die Daten aus den einzelnen Observatorien zusammenzuführen und der wissenschaftlichen bzw. nichtwissenschaftlichen Öffentlichkeit zugänglich zu machen. Hierzu wurden eine Reihe von Werkzeugen wie z.B. der hierarchischen Suche nach Stichworten oder durch Web-GIS- Funktionalitäten entwickelt, die es dem Benutzer erlauben, einen detaillierten Einblick in die verschiedenen Observatorien, Testgebiete und Messnetze zu erhalten. Datenvisualisierungswerkzeuge erlauben es, die Daten von verschiedenen Zeitreihen und Variablen zu selektieren, darzustellen, zu überprüfen und schließlich konform der gemeinsamen Datenpolitik auch herunter zu laden. Momentan sind zeitaktuelle Daten von mehr als 450 automatischen Stationen frei verfügbar.

Die Definition und Umsetzung geeigneter Schnittstellen und Standards für den Datenaustausch zwischen den lokalen Dateninfrastrukturen und TEODOOR ist die größte Herausforderung. In TERENO erfolgt dieser Austausch über standardisierte OGC-konforme Web- Dienste, die für jedes Observatorium aufgebaut und betrieben werden und welche die Daten und Metadaten für die interne und öffentliche Verwendung bereitstellen. Ein neu entwickelter OGC-konformer Sensor Observation Service (SOS) ist in der Lage, Rasterzeitreihendaten liefern zu können, und eröffnet damit eine Reihe erweiterter Funktionalitäten für die Veröffentlichung von Wetterradar- und Fernerkundungsdaten.

#### **Literaturverzeichnis**

- 52°North (2012): Sensor Web Community. 52°North - Initiative for Geospatial Open Source Software GmbH, <http://52north.org/communities/sensorweb/>, accessed 2012-08-30.
- Band, L., Ogden, F., Goodrich, D., Hooper, R., Kane, D., Lyons, B., McKnight, D., Miller, N., Williams, M., Potter, K., Scanlon, B., Pielke, R.A., Sr., Reckhow, K. (2005): Designing a network of hydrologic observatories as a community service. Technical Report Number 7, CUAHSI, pp. 44.
- Bogena, H.R., Herbst, M., Huisman, J.A., Rosenbaum, U., Weuthen, A., Vereecken, H. (2010): Potential of Wireless Sensor Networks for Measuring Soil Water Content Variability. *Vadose Zone Journal* (4): 1002-1013, doi: 10.2136/vzj2009.0173.
- Bogena, H.R., Schulz, K., Vereecken, H. (2006): TERENO: Towards a network of observatories in terrestrial environmental research. *Advances in Geosciences*: 109-114
- Bose, R., 2002. A conceptual framework for composing and managing scientific data lineage, 14th International Conference on Scientific and Statistical Database Management. IEEE Press, Piscataway, N. J., pp. 15-19.
- Botts, M. (Ed.), (2007): OpenGIS Sensor Model Language (SensorML) - Implementation Specification. DocNr. OGC-07-000, Ver 1.0.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 180.

- Botts, M., Percivall, G., Reed, C., Davidson, J. (2008): OGC Sensor Web enablement: Overview and high level architecture. In: Nittel, S., Labrinidis, A., Stefanidis, A. (Eds.), *Geosensor Networks. Lecture Notes in Computer Science*. Springer-Verlag Berlin, Berlin, pp. 175-190.
- Bronstert, A., Kneis, D., Bogena, H. (2009): Interactions and feedbacks in hydrological change: Relevance and possibilities of modelling. *Hydrologie und Wasserbewirtschaftung* (5): 289-304
- Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R. (2011a): New Generation Sensor Web Enablement. *Sensors* (3): 2652-2699, doi: doi:10.3390/s110302652.
- Bröring, A., Maué, P., Janowicz, K., Nüst, D., Malewski, C. (2011b): Semantically-Enabled Sensor Plug & Play for the Sensor Web. *Sensors* (8): 7568-7605
- Bröring, A., Stasch, C., Echterhoff, J. (Eds.) (2012): *Sensor Observation Service Interface Standard*. DocNr. OGC 12-006, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 163.
- Burt, T.P., Howden, N.J.K., Worrall, F., Whelan, M.J. (2008): Importance of long-term monitoring for detecting environmental change: lessons from a lowland river in South East England. *Biogeosciences*: 1529-1535
- Chen, N.C., Di, L.P., Yu, G.N., Gong, J.Y., Wei, Y.X. (2009): Use of ebRIM-based CSW with sensor observation services for registry and discovery of remote-sensing observations. *Computers & Geosciences* (2): 360-372, doi: 10.1016/j.cageo.2008.08.003.
- Cox, S. (Ed.), (2010): *Geographic Information: Observations and Measurements - OGC Abstract Specification Topic 20*. DocNr. OGC 10-004r3, ISO 19156, Ver 2.0.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 57.
- Cox, S. (Ed.), (2011): *Observations and Measurements - XML Implementation*. DocNr. OGC 10-025r1, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 76.
- Desilets, D., Zreda, M., Ferre, T.P.A. (2010): Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays. *Water Resources Research*, doi: W11505 10.1029/2009wr008726.
- ECJRC (Ed.), (2009): *INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119*. European Commission Joint Research Centre, pp. 74.
- FGDC (Ed.), (2000): *Content Standard for Digital Geospatial Metadata Workbook*. Federal Geographic Data Committee, Reston, Virginia, USA, pp. 126.
- Gray, J., Liu, D., Nieto-Santisteban, T.M., Szalay, A., DeWitt, D.J., Heber, G. (2005): Scientific data management in the coming decade. *SIGMOD Rec.* (4): 34-41, doi: doi:10.1145/1107499.1107503.
- Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I. (2008): A relational model for environmental and water resources data. *Water Resources Research* (5), doi: W05406 10.1029/2007wr006392.
- ISO (Ed.), (2003): *ISO/DIS 19115: Geographic information - Metadata*. DocNr. ICS 35.240.70, Ver ISO/DIS 19115. International Organization for Standardization, pp. 163.
- ISO (Ed.), (2009): *Information and documentation - The Dublin Core metadata element set*. DocNr. ISO 15836:2009. ISO, pp. 6.
- Jirka, S., Bröring, A., Kjeld, P., Maidens, J., Wytzisk, A. (2012): A Lightweight Approach for the Sensor Observation Service to Share Environmental Data across Europe. *Transactions in GIS* (3): 293-312
- Jirka, S., Bröring, A., Stasch, C. (2009): Discovery Mechanisms for the Sensor Web. *Sensors* (4): 2661-2681, doi: 10.3390/s90402661.
- Keller, M., Schimel, D.S., Hangrove, W.W., Hoffman, F.M. (2008): A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment* (282-284)
- kZen Labs (2012): Buddata ebXML Registry/Repository, <http://www.buddata-open.org/>, accessed 2012-08-30.
- Lesage, N. (Ed.), (2007): *OGC Cataloguing of ISO Metadata (CIM) - Using the ebRIM profile of CS-W*. DocNr. OGC 07-038, Ver 0.1.7. Open Geospatial Consortium, pp. 95.
- Lin, H. (2003): *Hydropedology: bridging disciplines, scales, and data*. *Vadoze Zone Journal*: 1-11
- Lin, H. (2010): Earth's Critical Zone and hydropedology: concepts, characteristics, and advances. *Hydrol. Earth Syst. Sci.*: 25-45
- McDonnell, J.J., Sivapalan, M., Vachè, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J., Weiler, M. (2007): Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resour. Res.* (7): W07301, doi: 10.1029/2006wr005467.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G. (1997): Nongeospatial metadata for the ecological sciences. *Ecological Applications* (1): 330-342, doi: 10.2307/2269427.

- Montgomery, J.L., Harmon, T., Haas, C.N., Hooper, R., Clesceri, N.L., Graham, W., Kaiser, W., Sanderson, A., Minsker, B., Schnoor, J., Brezonik, Patrick (2007): The WATERS Network: An Integrated Environmental Observatory Network for Water Research. *Environmental Science & Technology* (19): 6642-6647, doi: 10.1021/es072618f.
- Na, A., Priest, M. (Eds.) (2007): Sensor Observation Service. DocNr. OGC 06-009r6, Ver 1.0. Open Geospatial Consortium Inc., Wayland, MA, USA, pp. 104.
- Nebert, D., Whiteside, A., Vretanos, P. (Eds.) (2007): OpenGIS Catalogue Services Specification. DocNr. OGC 07-006r1, Ver 2.0.2, Corrigendum 2 Release. Open Geospatial Consortium, Wayland, MA, USA, pp. 218.
- Nisbet, E. (2007): Earth monitoring: Cinderella science. *Nature*: 789-790
- Nogueras-Iso, J., Zarazaga-Soria, F.J., Béjar, R., Álvarez, P.J., Muro-Medrano, P.R. (2005): OGC Catalog Services: a key element for the development of Spatial Data Infrastructures. *Computers & Geosciences* (2): 199-209, doi: 10.1016/j.cageo.2004.05.015.
- OASIS (Ed.), (2002): OASIS/ebXML Registry Information Model. Ver 2.0. The Organization for the Advancement of Structured Information 1640 Standards [OASIS], pp. 60.
- OGC (2012): OGC Standards and Supporting Documents. Open Geospatial Consortium, <http://www.opengeospatial.org/standards>, accessed 2012-08-20.
- Parr, T.W., Sier, A.R.J., Battarbee, R.W., Mackay, A., Burgess, J. (2003): Detecting environmental change: science and society, Perspectives on long-term research and monitoring in the 21st century. *Science of the Total Environment*: 1-8, doi: 10.1016/s0048-9697(03)00257-2.
- Plone Foundation (2004): Plone, <http://www.plone.org>, accessed 2012-08-30.
- Reid, W.V., Bréchinac, C., Tseh Lee, Y. (2009): Earth System Research Priorities. *Science* (5938): 245-245
- Richter, D.d., Mobley, M.L. (2009): Monitoring Earth's Critical Zone. *Science* (5956): 1067-1068
- Simonis, I., Echterhoff, J. (Eds.) (2011): OGC Sensor Planning Service Implementation Standard. DocNr. OGC 09-000, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 248.
- Simonis, I., Wytzisk, A. (Eds.) (2003): Web Notification Service. DocNr. OGC 03-008r2, Ver 0.1.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 46.
- Sorg, J. (2012): Entwurf, Implementierung und Anwendung eines OGC-konformen Sensor Observation Service für flächenbezogene Rasterzeitreihendaten. Master Thesis, Fernuniversität Hagen, Hagen, Germany, 96 pp.
- Tarboton, D.G., Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., Goodall, J., Valentine, D., Whitenack, T. (2009): Development of a Community Hydrologic Information System. 18th World Imacs Congress and Modsim09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, 988-994 pp.
- TERENO (2010): TERENO Data policy, [http://teodoor.icg.kfa-juelich.de/downloads/TERENO\\_Data\\_policy.pdf/](http://teodoor.icg.kfa-juelich.de/downloads/TERENO_Data_policy.pdf/), accessed 2012-08-20.
- Tomasic, A., Simon, E. (1997): Improving access to environmental data using context information. *SIGMOD Rec.* (1): 11-15, doi: 10.1145/248603.248606.
- Vanderborght, J., Graf, A., Steenpass, C., Scharnagl, B., Prolingheuer, N., Herbst, M., Franssen, H.-J.H., Vereecken, H. (2010): Within-Field Variability of Bare Soil Evaporation Derived from Eddy Covariance Measurements. *Vadose Zone Journal* (4): 943-954, doi: 10.2136/vzj2009.0159.
- Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuss, R., Puetz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J.C., Papen, H., Priesack, E., Schmid, H.P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., Vereecken, H. (2011): A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone Journal* (3): 955-973, doi: 10.2136/vzj2010.0139.
- Zoback, M.L. (2001): Grand challenges in earth and environmental sciences: sciences, stewardship, and service for the twenty-first century. *GSA Today*: 41-47

## **Auf kreativen Wegen von Daten zum Wissen am Beispiel medizinischer Forschungsdaten**

Lars Müller, Thomas Wetzel, Hans-Christoph Hobohm

FH-Potsdam, FB Informationswissenschaften

### **Zusammenfassung**

DataCreativityTools ist ein BMBF-Projekt mit dem Ziel, das Innovationspotential medizinischer Forschungsdaten besser auszuschöpfen, indem ein Creativity Support Tool entwickelt wird. Es wird ein Ansatz verfolgt, bei dem Hypothesenfindung im datenorientierten Forschungsprozess als kreativer Prozess modelliert wird. Durch die Implementierung dieses Prozesses in einer Kreativität fördernden Informationsumgebung sollen die Anwender darin unterstützt werden, neue Forschungsfragen und -hypothesen zu entwickeln.

### **Abstract**

DataCreativity is a research & development project aiming to implement prototype creativity support tools for hypotheses generation. Hypotheses generation is modeled as creative process initiated by biomedical research data.

### **1 Einleitung: Das Problem ist das Problem**

Im Jahr 1870 publizierte Eugen Semmer im Archiv für Pathologische Anatomie und Physiologie über seine Tierversuche an Pferden. Auf der Suche nach dem Nachweis für den Milzbranderreger injizierte er den Tieren zahlreiche Pilzsporen, u.a. auch solche mit heilender Wirkung, die erst Jahre später als Grundlage für Penicillin in die Medizingeschichte eingingen. Seine Versuchsreihe konnte er erst erfolgreich abschließen, als es ihm gelang, ein Fohlen mit dem Milzbranderreger zu infizieren. Die gesunden Tiere waren für ihn eine Störung bei der Lösung seines Problems.<sup>1</sup> (Semmer 1870) Eine andere Problemstellung hätte ihn zum Entdecker des Penicillins machen können.

Wir vertreten die These, dass in bereits vorhandenen Daten Lösungen für wichtige Probleme in greifbarer Nähe sind, sie aber nicht gefunden werden, wenn am „falschen“ Problem geforscht wird.

---

<sup>1</sup> Den Hinweis auf das genannte Beispiel verdanken wir Cropley und Cropley (2009: 71), die die Geschichte allerdings etwas anders darstellen.

## 2 Daten, Kreativität, Innovation

Integration und semantische Anreicherung von Forschungsdaten schaffen neue Datenräume. Dadurch entsteht eine Datenbasis, die so kein einzelnes Experiment bzw. keine Erhebung hervorbringen könnte. Es sind jedoch nicht allein die Akkumulation ungeheurer Datenmengen und deren automatische Verarbeitung, die zu Neuerungen führen, sondern eben auch die vernetzende Infrastruktur, die zunehmend besser semantische Bezüge abbilden kann und neben der Datenbasis auch die Erprobung neuer wissenschaftlicher Methoden ermöglicht.

Die quantitative Zunahme der Forschungsdaten, deren Vernetzung sowie der Ausbau virtueller Forschungsumgebungen (Arbeitsgruppe „Virtuelle Forschungsumgebungen“ 2011), kurz: „e-Science“, beinhalten ein großes Innovationspotential.<sup>2</sup> „Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis“ (Allianz der deutschen Wissenschaftsorganisationen 2010: [2]) und sollen deshalb in unserem Projekt den Ausgangspunkt für die Schaffung von Innovationen bilden. „Innovation involves the introduction of effective novelty into a functioning system. The first step is the generation of the effective novelty; this is the task of creativity.“ (Cropley und Cropley 2009: 2) Um das Innovationspotential der e-Science besser auszuschöpfen, bedienen wir uns deshalb bei der Kreativitätsforschung.

Eine gewisse Menge Ideen zu sammeln oder zu erzeugen, ist nicht ausreichend, um von wirklicher Kreativität zu sprechen. „Creativity has two components: generation of novelty and evaluation of the relevance and effectiveness of the novelty.“ (Cropley und Cropley 2009: 27) Die Neuerung muss folglich nicht nur für einen selbst neu sein, sondern auch für andere und sie muss sich durch Anerkennung bzw. Anwendung bewähren, darf also auch nicht beliebig sein. (Cropley und Cropley 2009: 29ff.) Cropley & Cropley beschreiben Kreativität als einen siebenstufigen Prozess, der alle Phasen von den Voraussetzungen bis zur Einführung bzw. Anerkennung einer kreativen Entwicklung umfasst:

1. *Preparation* (Bestand und Ausbau der Wissensbasis),
2. *Activation* (Problemerkennung),
3. *Generation* (Ideenfindung),
4. *Illumination* (Fokussierung auf Lösungsmöglichkeit, das kreative Produkt),

---

<sup>2</sup> Zu den Begriffen Forschungsdaten und e-Science siehe: Büttner et al. (2011: 13-17).



5. *Verification* (Kritische Prüfung und Auswahl des Produkts),
6. *Communication* (Kommunizierung der Lösung/des Produkts),
7. *Validation* (Beurteilung und Anerkennung des Produkts durch andere).  
(Cropley und Cropley 2009: 87ff.)

Neben möglichst umfassendem Fachwissen auf dem Gebiet, in dem die kreative Entwicklung stattfindet, erfordert der kreative Prozess je nach Phase wechselnd sowohl divergentes als auch konvergentes Denken. Divergentes Denken dient der Erzeugung von Variabilität, während konvergentes Denken stets zur kritischen Prüfung der neuen Gedanken erforderlich ist und dem Prozess eine sinnvolle Richtung gibt. Beide Denkweisen müssen im kreativen Prozess zur Anwendung kommen und dürfen sich gegenseitig nicht blockieren. (Cropley und Cropley 2009: 79ff.)

Eine Schlüsselfunktion im kreativen Prozess kommt dem Problemfinden bzw. -definieren zu. In der Wissenschaft werden mit den Hypothesen die Gegenstands- und Problemfelder bestimmt, die untersucht werden. Schon Ludwig Fleck, Pionier der soziologischen Wissenschaftsforschung, berichtete von der Schwierigkeit, bei seiner Forschung das Problem herauszuarbeiten: Eine Versuchsreihe zu Bakterienkolonien „verlangte erst viele tastende Versuche zur Methodenwahl und viele Formulierungsproben des Problems. Nicht einmal das konnte klar gesagt werden, dass ein wirkliches Problem vorliege.“ (Fleck 1980, S.116) Die Wahl eines Forschungsproblems determiniert das Forschungsergebnis. Ob jedoch ein Problem erkannt wird und welches schließlich zur Erzeugung wissenschaftlicher Tatsachen weiterverfolgt wird, unterliegt nicht allein dem Material bzw. den Daten, sondern auch der persönlichen Disposition der Forscher/innen, ihrer Fachtradition und der historischen Epoche. (Fleck 1980, S. 117f., S.123f., S. 137) Da die Problemfindung (das Finden und Formulieren „guter“ Probleme oder, noch zugespitzter, das Erfinden relevanter Probleme) nach Ansicht der Kreativitätsforschung der essentielle Kern für die Entwicklung von Neuem ist (Cropley und Cropley 2009: 253f.), rückt sie in unseren Fokus, um aus Forschungsdaten Innovationen zu generieren.

Lässt sich der kreative Prozess, die Kreativität, gezielt fördern? Bei dieser Frage denkt man zunächst an den Einsatz von Kreativitätstechniken und Problemlösungsmethoden, um innovative Ideen zu generieren.

Brainstorming ist wohl das prominenteste Beispiel für eine sogenannte Kreativitätstechnik. Allerdings funktioniert das Brainstorming nicht sonderlich gut und



im besten Falle kann es lediglich dazu dienen, Gruppenprozesse positiv zu beeinflussen und bereits vorhandenes zu sammeln und zu fixieren. (Runco 2007: 350f.) In der Vergangenheit richteten sich viele Anstrengungen darauf, Problemlösungsmethoden zu entwickeln. Es gibt zahlreiche Taktiken, die die Entwicklung kreativer Ideen begünstigen sollen.<sup>3</sup> (Runco 2007: 319f.) Eine einfache Technik zur Erzeugung von Kreativität gibt es nicht. Wichtige Voraussetzungen liegen in den persönlichen Fähigkeiten der Forscherinnen und Forscher, ihrer Ausbildung und Bereitschaft zu kreativem Arbeiten. Die erfolgversprechendste Art, Kreativität zu fördern, ist es, ein anregendes Umfeld für kreatives Arbeiten zu schaffen und Kreativität hemmende Einflüsse möglichst weit zu reduzieren (Runco 2007: 371ff.) sowie Teilnahme an einem Training zum kreativen Arbeiten. (Cropley und Cropley 2009: 224ff.)

Wenn wir von Creativity Support Tools sprechen, meinen wir folglich nicht ein Brainstormingprogramm auf dem Computer oder eine vergleichbare Software, die Ideefindung oder eine andere Phase der Kreativität mehr begleitet als unterstützt. Computer gestützter Creativity Support wird eher erzielt durch die Berücksichtigung von Designprinzipien, als durch die Implementierung einzelner Techniken. Human-Computer-Interface (HCI) Designer haben Gestaltungsprinzipien ausgearbeitet, um Softwareanwender in kreativen Arbeitsprojekten optimal zu unterstützen. Sie entwickeln „creativity support tools, which enable users to explore, discover, imagine, innovate, compose, and collaborate.“ (Shneiderman 2009: 1) Diese Prinzipien legen wir als Richtschnur für unsere Entwicklung an.

### **3 Das Projekt DataCreativityTools for Innovation and Research**

Im Projekt *DataCreativityTools for Innovation and Research* (DCT) an der Fachhochschule Potsdam werden in enger Kooperation mit dem *Open European Nephrology Science Center* (OpEN.SC) der Charité – Universitätsmedizin Berlin praktische Lösungen entwickelt, um das in den Forschungsdaten liegende Erkenntnispotential besser zu nutzen.

Das DFG-Projekt OpEN.SC hat die Implementierung, Evaluation und Bereitstellung eines offenen, europäischen Leistungszentrums für Forschungsinformation für Klinik und Pathologie, Lehre und Forschung in Nephrologie und nephrologischer

---

<sup>3</sup> Die Empfehlungen reichen von „Perspektivwechsel“ oder „Suche nach Analogien und Gegensätzen“ über sequentielle Konzepte zur Problemlösung (Creative Problem Solving, CPS) bis zur TRIZ-Technik, die 40 Prinzipien zum Erfinden innovativer Lösungen umfasst.

Transplantationsmedizin zum Gegenstand. Die Daten innerhalb des OpEN.SC Portals bilden die Grundlage für die Initiierung, Durchführung, Auswertung und Publikation von experimentellen und klinischen Studien, Diskussionen und Kommunikation. (Schrader et al. 2007)

Ziel des DCT-Projekts ist es, auf informationswissenschaftlicher Grundlage ein prototypisches Werkzeug zu entwickeln, dass die Generierung neuer Forschungshypothesen aus diesen medizinischen Daten unterstützt.

### **3.1 OpEN.SC-Daten**

Ein wesentlicher Arbeitsschritt innerhalb des OpEN.SC Workflows besteht in der Importierung medizinischer Daten aus verschiedenen Kliniken und Instituten der Charité. Die Daten zeichnen sich vor allem durch ihre Heterogenität aus. Sie beinhalten beispielsweise natürlichsprachliche Texte (Befunde, Diagnosen), numerische Daten (Laborwerte) aber auch Bilddaten wie Virtual Slides. Virtual Slides sind digitalisierte Aufnahmen von mikroskopischen Objektträgern, die mit entsprechender Software „virtuell“ am Bildschirm betrachtet werden können. Kleine Domänen wie die Nephropathologie verfügen an ihrem jeweiligen Standort über einen vergleichsweise geringen Datenbestand. Erst die Zusammenführung in einer zentralen Datenbasis intensiviert die Möglichkeiten für unterschiedlichste Forschungsfragestellungen. Der Importierungsvorgang ist als mehrstufiges Verfahren angelegt. In einem ersten Schritt wird eine grobe Qualitätsanalyse durchgeführt. Falsche und fehlerhafte Datensätze werden hierbei eliminiert. Anschließend werden personenbezogene Inhalte verfremdet und somit die Daten anonymisiert. Um die Speicherung unterschiedlichster Daten zu gewährleisten, werden die Daten anschließend in das RDF Format überführt. Dabei werden die jeweiligen Attributnamen mit einem OpEN.SC Standardvokabular versehen. Auf diesem Weg entsteht durch die Zusammenführung der Daten aus den verschiedenen Zentren und Instituten eine beträchtliche Datenbasis welche für diverse Forschungsfragestellungen zur Verfügung steht. Derzeit wird die Schaffung weiterer Schnittstellen zu externen Systemen sowie die Integration zusätzlicher medizinischer Standards diskutiert.

### 3.2 Nutzung von Daten in der Anwendungsdomäne Pathologie

Im Rahmen unseres Projekts haben wir uns ein Bild davon gemacht, wie unsere potentielle Anwendergruppe mit Daten umgeht. Unsere Befragungsergebnisse gestatten einige Aussagen hierzu.<sup>4</sup>

Den Forscherinnen und Forschern stehen grundsätzlich Daten zur Verfügung, und sie sagen von sich, sie auf unterschiedlichen Wegen zu beschaffen. Dafür werden Quellen mit guter Qualität und Verfügbarkeit bevorzugt (Datenzentren und eigene Erhebungen vor Patientenakten und Daten von Kollegen). Daten werden für wissenschaftliche Arbeit genutzt, überwiegend, um eine spezifische Forschungsfrage oder neue Ideen zu verfolgen. Sie werden vor allem dort eingesetzt, wo wissenschaftliche Arbeit solide untermauert werden muss: in Publikationen, Projektanträgen und Patenten. Spontane, ungerichtete Exploration von Daten findet kaum statt und bildet nicht den Ausgangspunkt für die Entwicklung neuer (Forschungs-)Ideen.

Die Mediziner haben offenbar eine hohe Datenaffinität, nutzen Daten aber kaum zur unmittelbaren Suche nach Forschungsfragen und -hypothesen. Für unseren Anwendungsfall können wir daher feststellen, dass das Potential für datenorientierte Forschung vorhanden ist, aber nicht ausgeschöpft wird. Das beschriebene Anwendungsfeld ist für die Implementierung und Erprobung von DCT gut geeignet.

## 4 Der datenorientierte empirische Forschungsprozess

Kell & Oliver fragten bereits vor einigen Jahren: „Here is the evidence, now what is the hypothesis?“ Sie verwiesen darauf, dass es auch in der Medizin bereits Felder gibt, die genuin Daten getriebene Forschung betreiben: „hypotheses are the result of the epidemiological study of interest, not its starting point.“ (Hv. im Original) (Kell und Oliver 2004: 101)

Unser Anliegen ist es, einen Weg aufzuzeigen, um datenorientierte Forschungsprozesse auf bislang Hypothesen orientierte Wissenschaften auszudehnen. Die Methodologie empirischer Forschung kennt seit langem explorative Verfahren zur Findung von Hypothesen. (Bortz und Döring 2006: 358ff.) Die entscheidende Neuerung durch den Trend zur e-Science ist die quantitative Zunahme der Daten und deren schnellere wie auch leichtere Verfügbarkeit und Verarbeitung, z.B. in Visualisierungen.

---

<sup>4</sup> Wir haben Leitfaden gestützte Interviews und eine Onlinebefragung durchgeführt.

Thomas S. Kuhn hat den Vorgang wissenschaftlicher Entdeckungen wie folgt beschrieben:

*„Die Entdeckung beginnt mit dem Bewußtwerden einer Anomalie, das heißt mit der Erkenntnis, daß die Natur in irgendeiner Weise die von einem Paradigma erzeugten, die normale Wissenschaft beherrschenden Erwartungen nicht erfüllt hat. Sie geht dann weiter mit einer mehr oder weniger ausgedehnten Erforschung des Bereichs der Anomalie und findet erst einen Abschluß, nachdem die Paradigmatheorie so berichtigt worden ist, daß das Anomale zum Erwarteten wird.“ (Kuhn 2009: 65f.)*

Für datenorientierte Forschung heißt das, dass die Datenexploration den Startpunkt einer Forschung bilden muss, denn nur so kommt der Impuls für die Entdeckung (der Anomalie) unmittelbar aus den Daten selbst. In dem Spannungsfeld zwischen Paradigma und Anomalie werden schließlich neue Forschungshypothesen generiert. Wenn dann neue Beobachtungen in Übereinstimmung mit dem Paradigma gebracht wurden, ist die Anomalie aufgehoben. Handelt es sich um „normale Wissenschaft“ (Kuhn 2009: 47), geschieht dies durch Verfeinerung und Anpassung der Theorie innerhalb des bestehenden Paradigmas. Kann die Anomalie nicht ohne weiteres mit bewährten Methoden aufgehoben werden, besteht die Chance, neue Probleme zu finden.

Aus Perspektive eines kreativen Prozesses stellt sich dieser Vorgang folgendermaßen dar:

*„Problem finding processes are initiated as attempts are made to organize and utilize information during the acquisition of further domain-specific content. What is implied here is that with an effective knowledge base the individual is capable of recognizing when new information does not fit into the existing cognitive structures. (...) Problem finding, then, results from an effort to utilize both specific and general problem solving procedures as an attempt is made to integrate new data, experience, or information into an organized memory structure.“ (Hoover und Feldhusen 1994: 213f.)*

Problemfindung entwickelt sich also aus gescheiterten Problemlösungsversuchen.

Da es das Ziel der Forschung ist, relevante Probleme zu lösen, ist es nicht ausreichend, irgendetwas Neues zu finden, sondern es muss auch im bestehenden fachlichen Forschungsumfeld eingeordnet werden und dort sowohl als relevanter als auch neuer Gegenstand Bestand haben. Hierfür ist die Verbindung der Daten mit dem zur Verfügung stehenden Wissen erforderlich. Diese Verbindung wird durch Informationsprozesse realisiert.

Das persönliche Wissen der Forscher bildet zusammen mit isoliertem Faktenwissen und publiziertem wissenschaftlichen Wissen die Grundlage. Wenn wirklich neue Hypothesen gefunden werden sollen, kann der Informationsprozess aber nicht zu Antworten führen. Er muss in das Feld der „umstrittenen Tatsachen“ (Latour 2004) führen. Dieses Feld ist die aktuelle „Zeitschriftenwissenschaft“. (Fleck 1980: 156f.) In diesem Feld muss sich eine neue Forschungshypothese als kreatives Produkt bewähren. Tut sie das, kann sie in einer neuen Forschung der empirischen Prüfung unterzogen werden und zur Lösung von wesentlichen Problemen beitragen.

Im klassischen empirischen Forschungsprozess schließt sich der Hypothese die Datensammlung an. Dieser Vorgang wird auch in der e-Science nicht überflüssig werden. Er wird jedoch ergänzt um die initiale Datenerkundung zur Hypothesenbildung. Hypothesenorientierte Forschung wird in diesem Sinne durch datenorientierte Forschung erweitert, nicht ersetzt. Beide Formen ergänzen sich gegenseitig und werden im Wechsel durchgeführt. Vgl. Kell und Oliver (2004: 103)

## **5 Entwicklung der DCT**

### **5.1 DCT-Entwicklungsansatz**

Ziel unserer Entwicklung ist es, mit den DCT eine Informationsumgebung zu schaffen, die nach den oben erwähnten HCI-Designprinzipien für Creativity-Support gestaltet ist (s.o., Shneiderman 2009), um die Kreativität der Anwender anzuregen. Die Forschungsdaten für unseren prototypischen Anwendungsfall werden dafür semantisch angereichert und mit externen Wissensbeständen in Beziehung gesetzt. Dadurch wird eine Visualisierung der Daten in einem komplexen Bedeutungsnetz ermöglicht. Die DCT leiten die Forscher bei der Datenerkundung durch einen Prozess, der anhand von Erkenntnissen aus der Kreativitätsforschung modelliert wird und an dessen Ende neue Forschungsfragen stehen.

Bei der DCT-Entwicklung liegt der Fokus auf der Phase vor der Hypothesenbildung. Die Bedeutung des Problemfindens für kreative Arbeit haben wir oben bereits betont. Sie stellt uns jedoch vor eine besondere Herausforderung: „Scientists are generally agreed on how this [hypothesis testing, L.M.] is to be done, and the basic rule is that testing procedures be public and replicable. Hypothesis generation is a much more private affair and not well understood.“ (Nickerson 1999: 395)<sup>5</sup>

---

<sup>5</sup> Vgl. auch Bortz und Döring (2006: 358).

Die Abarbeitung von wohldefinierten Problemen bzw. Hypothesenprüfungen stellt auch für Computer keine Schwierigkeit dar. Problemfindung allerdings ist ein Vorgang, der bislang noch nicht automatisiert werden konnte. (Runco 2007: 18)

## 5.2 Prozessmodell

Is Grundlage für unsere Entwicklung haben wir ein Prozessmodell entwickelt, das auf den beschriebenen Kreativitäts- und Informationsvorgängen basiert.

(Abb. 1) Kerngedanke des Modells ist es, prozessschrittspezifisch divergentes und konvergentes Denken anzuregen.

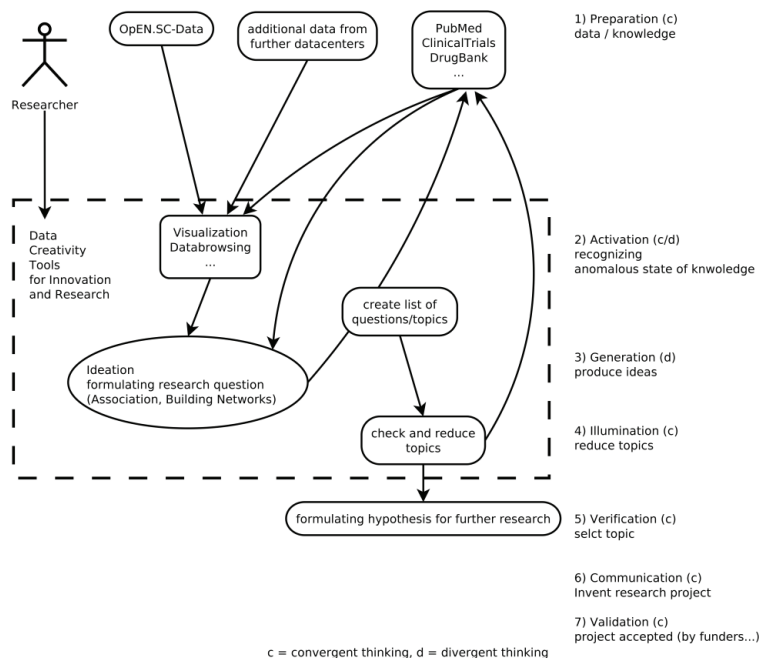


Abbildung 1: DCT-Prozessmodell

Dabei betrachten wir das Finden von Problemen für sich genommen als kreativen Prozess. Das Modell stützt sich auf das „extended Phase Model“ von Cropley & Cropley (2009: 89), das den kreativen Prozess als strukturierte Abfolge von Schritten darstellt, die durch konvergentes oder divergentes Denken geprägt sind. (Cropley und Cropley 2009: 80f.)

Jedwede Neuerung fußt auf der Wissensbasis, die einem kreativen Individuum zur Verfügung steht. Für DCT sind dies das persönliche Wissen der Forscher/innen und das ihnen mittels Informationssystemen zur Verfügung stehende ergänzende

Wissen. Die Herstellung der Wissensbasis steht für die erste Phase im Kreativitätsprozess. In der zweiten Phase „Activation“ wird ein Problem erkannt. Hierfür bieten DCT verschiedene Möglichkeiten, Forschungsdaten zu explorieren, um divergentes Denken anzuregen. Um überhaupt eines Problems gewahr zu werden, bedarf es stetiger Bezugnahme auf die Wissensbasis. Es entsteht der „Anomalous State of Knowledge“ (Belkin 2006). Aus diesem Zustand heraus werden unspezifische Informationssuchen ausgelöst, um die „Wissensordnung“ wieder herzustellen. An dieser Stelle setzt im kreativen Prozess der Informationsprozess ein, in dem versucht wird, mithilfe zusätzlicher Information den unausgeglichene Wissenszustand wieder herzustellen (s.o.). Der unsichere Zustand, das „unklare anfängliche Schauen“ (Fleck 1980: 121), soll von DCT zunächst verstärkt werden. Anwender erhalten verschiedene Möglichkeiten (Visualisierungen, Browsen durch die Datenfelder und Daten), die Daten zu erkunden. Indem die Bildschirmaktivitäten als Suchanfragen konzeptualisiert werden, werden begleitend zur Datenerkundung kontextrelevante Information angeboten, die die Wissensbasis erweitern und einen weiten Fokus anregen sollen. Dies soll gleichzeitig die Ideenfindung durch neues Wissen oder die Arbeit von anderen anregen und die kritische Prüfung durch Zugriff auf bekannte Fakten etc. unterstützen.

Jeder Teilinformationsprozess in den DCT hat zum Ziel, ein mögliches Problem zu erkennen. Insgesamt sollen in diesen Phasen potentielle Forschungsfragen erzeugt werden. Indem die Anwenderaktivitäten protokolliert werden, wird es möglich, die potentiellen Forschungsfragen in einem weiteren Schritt zu evaluieren und mit dem gleichen Instrumentarium der Informationsrecherche zu bewerten. Im Idealfall ist bei der Bewertung der Daten alles Wissen bekannt und die Suche in der aktuellen Forschung bleibt ohne Ergebnis bzw. liefert keine Antworten sondern umstrittene Positionen.

Erfolgreicher Abschluss des DCT-Prozesses ist die Formulierung einer Forschungshypothese, die ihre wesentlichen Impulse der Datenexploration verdankt. Außerhalb der DCT erfolgt die Vollendung des kreativen Prozesses. Erst wenn das kreative Produkt (die Hypothese) kommuniziert wurde und von der Fachcommunity anerkannt wird, kann von einer wirklichen Neuerung gesprochen werden. Ein geeigneter Indikator zur Feststellung eines erfolgreichen Kreativitätsprozesses wäre bspw. die Bewilligung von Finanzmitteln für ein neues Forschungsprojekt.

### 5.3 CT-Systemmodell

Wie bereits an unserem Prozessmodell deutlich geworden ist, arbeiten wir daran, eine Arbeits- und Wissensumgebung zu schaffen, die Kreativität begünstigt. Wir können uns dafür an den Designprinzipien für Creativity Support Tools orientieren. (Shneiderman 2009: 4f.). Creativity Support leisten wir durch den Aufbau einer Informationsumgebung, die den kreativen Vorgang der Hypothesenfindung unterstützt.

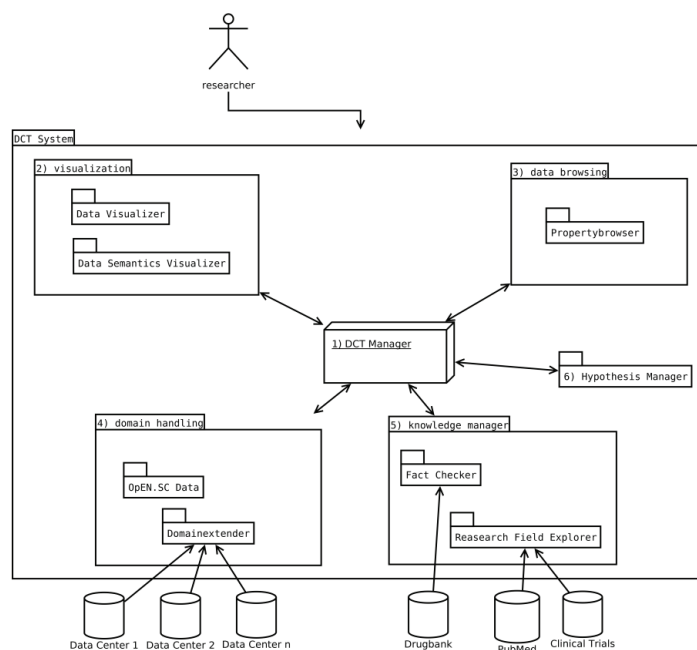


Abbildung 2: DCT-Systemmodell

Hierfür werden Module zusammengestellt, die Forschungsdatenexploration und Informationsprozesse im Semantic Web ermöglichen. (Abb. 2) Die OpEN.SC-Daten bilden den Ausgangspunkt für die Anwendung (Domain Handling). Ein Visualisierungsmodul und der Metadatenbrowser ermöglichen Exploration von Metadaten und Daten. Der Knowledge manager ist eine Information-Retrieval-Schnittstelle zu relevanten Quellen im Semantic Web. Der Hypothesenmanager protokolliert den Arbeitsprozess und unterstützt die Bewertung der Suchanfragen und Ideen. Die Entwicklung des Komponentenmodells ist die Grundlage für unsere weiteren Entwicklungsschritte.



## 6 Fazit

Der verfolgte Ansatz, eine Informationsumgebung zu schaffen, die sich positiv auf die Kreativität bei der Problemfindung und Hypothesenbildung der Anwender/innen auswirken soll, erschwert die Bewertung der Entwicklung hinsichtlich ihres Effekts. In Problemlösungsumgebungen lassen sich standardisierte Testläufe durchführen, die messbare Ergebnisse liefern. Anders in unserem Fall: Erfolgreicher Einsatz der DCT lässt sich nur mit qualitativen Untersuchungstechniken beurteilen und kann zum jetzigen Zeitpunkt noch nicht systematisch evaluiert werden. Zudem wirft unser Ansatz die Schwierigkeit auf, dass die Auswahl der Daten- und Informationsquellen ganz wesentlicher Bestandteil der Konfiguration ist. Auf andere Domänen lässt sich somit zwar unser Modell übertragen, eine technische Implementierung erfordert aber die intensive Auseinandersetzung mit den jeweiligen relevanten Datenzentren und Informationsinfrastrukturen. Selbst innerhalb einzelner wissenschaftlicher Teildisziplinen ist es heute noch nicht absehbar, dass Verschiedene Datenzentren automatisch, quasi „on-the-fly“, recherchiert und in Browsing- und Visualisierungstools integriert werden können.

Es ist uns unserer Auffassung nach dennoch gelungen, ein schlüssiges Modell zu entwickeln, dass die Hypothesenbildung in einer Informationsumgebung schematisch abbildet. Der Schlüssel hierzu ist das Konzept einer Kreativität begünstigenden und anregenden Informationsumgebung.

Die Koppelung von Kreativitätsprozess und Informationsprozess bietet für die Informationswissenschaft noch Potential ausgebaut zu werden. Alle Fragen rund um die Entstehung wissenschaftlicher Tatsachen in e-Sciences lohnen von den Informationswissenschaften weiter vertieft zu werden.

Ob Kreativität tatsächlich verstärkt (*enhanced*) werden kann, ist eine nicht abschließend geklärte Frage. Nickerson beantwortete sie für sich positiv und begreift jegliche Anstrengungen in diese Richtung als Chance, mehr darüber herauszufinden, wie dieses Ziel erfolgreich verfolgt werden kann. (Nickerson 1999: 407f.) Dem schließen wir uns an.

## Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten.  
[http://www.allianzinitiative.de/fileadmin/user\\_upload/Home/Video/Grunds%C3%A4tze%20Umgang%20mit%20Forschungsdaten.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/Home/Video/Grunds%C3%A4tze%20Umgang%20mit%20Forschungsdaten.pdf).
- Arbeitsgruppe „Virtuelle Forschungsumgebungen“. 2011. Definition Virtuelle Forschungsumgebung. Arbeitsgruppe „Virtuelle Forschungsumgebungen“ in der Schwerpunktinitiative „Digitale Information“ in der Allianz der deutschen Wissenschaftsorganisationen.  
[http://www.allianzinitiative.de/fileadmin/user\\_upload/2011\\_VRE\\_Definition.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/2011_VRE_Definition.pdf).
- Belkin, N.J. 2006. Anomalous State of Knowledge. In *Theories of Information Behavior*. ASIST monograph series, 2nd. Ed, Hrsg. K.E Fisher, S. Erdelez und L.E.F McKechnie, 44-48. Medford, NJ: Information Today.
- Bortz, Jürgen, und Nicola Döring. 2006. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Aufl. Berlin...: Springer.
- Büttner, Stephan, Hans-Christoph Hobohm, und Lars Müller. 2011. Research Data Management. In *Handbuch Forschungsdatenmanagement*, Hrsg. Stephan Büttner, Hans-Christoph Hobohm und Lars Müller, 13-24. Bad Honnef: Bock + Herchen.  
[http://opus.kobv.de/fhpotsdam/volltexte/2011/225/pdf/1.1\\_Research\\_Data\\_Management.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/225/pdf/1.1_Research_Data_Management.pdf)
- Cropley, Arthur, und David Cropley. 2009. *Fostering creativity. A diagnostic approach for higher education and organizations*. Cresskill, NJ: Hampton Press.
- Fleck, Ludwik. 1980. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*, Frankfurt am Main: Suhrkamp.
- Hoover, Steven M, und John F. Feldhusen. 1994. Scientific Problem Solving and Problem Finding: A Theoretical Model. In *Problem finding, problem solving, and creativity*. Creativity research, Hrsg. Mark A. Runco, 201-219. Norwood N.J: Ablex Pub. Corp.
- Kell, Douglas B, und Stephen G. Oliver. 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* 26, 99-105. doi: 10.1002/bies.10385
- Kuhn, Thomas S. 2009. *Die Struktur wissenschaftlicher Revolutionen*, Bd. 25. 2, rev. und um das Postskriptum von 1969 erg. Aufl, [Nachdr.]. Frankfurt am Main: Suhrkamp.
- Latour, Bruno. 2004. Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry* 30, 225-248.
- Nickerson, Raymond S. 1999. Enhancing Creativity. In *Handbook of creativity*, Hrsg. Robert J. Sternberg, 392-430. Cambridge: Cambridge Univ. Press.
- Runco, Mark A. 2007. *Creativity. Theories and themes ; research, development, and practice*: Elsevier.
- Schrader, T, B. Rudolph, M. Dietel, M. Beil, T. Schaaf, D. Schmidt, und G. Lindemann. 2007. The Open European Nephrology Science Centre (OpEN.SC) - an information service center for kidney diseases and transplantation. *PATHOLOGY RESEARCH AND PRACTICE* 203, 314.
- Semmer, Eugen. 1870. Resultate der Injectionen von Pilzsporen und Pilzhefen in's Blut der Thiere. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* 50, 158-160.
- Shneiderman, Ben. 2009. Creativity Support Tools: A Grand Challenge for HCI Researchers. In *Engineering the User Interface. From Research to Practice*, Hrsg. Manuel Ortega, Crescencio Bravo und Miguel Redondo, 1-9. London: Springer-Verlag London.

---

## **Linked Data als Infrastruktur zur Integration von Forschungsdaten und Publikationen**

Dominique Ritze, Kai Eckert

Universitätsbibliothek Mannheim

### **Zusammenfassung**

Die fehlende Integration von Publikationen und Forschungsdaten erschwert zum einen die wissenschaftliche Recherche und behindert zum anderen sowohl die Nachweisbarkeit als auch die Reproduzierbarkeit von Ergebnissen. Die Verknüpfungen zwischen Forschungsdaten und Publikationen werden u.a. derzeit im DFG-geförderten Projekt InFoLiS von der UB Mannheim in Kooperation mit der GESIS und dem Lehrstuhl für Künstliche Intelligenz der Universität Mannheim erstellt.

Die so gefundenen Verknüpfungen müssen in geeigneter Weise zur Verfügung gestellt werden, so dass sie in den vorhandenen Recherchesystemen genutzt werden können. Dazu bietet sich eine Repräsentation als Linked Data an.

In diesem Paper beschreiben wir die Möglichkeiten, solcherart zur Verfügung gestellte Daten in das eigene Recherchesystem zu integrieren. Wir unterscheiden server- und clientseitige Anreicherungen und zeigen die jeweiligen Vor- und Nachteile. Dabei gehen wir insbesondere auf die clientseitige Anreicherung ein, die minimal-invasiv umgesetzt werden kann und nur einen geringen Aufwand bedeutet. Des Weiteren zeigen wir, dass sich auch damit alle Anforderungen, die sich in unserem konkreten Anwendungsfall ergeben, erfüllen lassen.

### **Abstract**

The missing integration of publication and research data complicates the scientific research. Especially the traceability as well as the reproducibility of results is hindered. Within the InFoLiS project, funded by the DFG, the Mannheim University Library, the Chair of Artificial Intelligence at the University of Mannheim and the GESIS face the connection between publication and research data.

After the connections have been detected, they need to be presented within the research systems in a suitable way. For this purpose, Linked Data offers an appropriate representation.

In this paper, we describe the possibilities to integrate such data in own research systems. We divide these enrichment methods into server- and client-side approaches and show their advantages as well as disadvantages. Afterwards, we especially focus on client-side enrichments. Such an approach can be implemented in a minimal-invasive way requiring only a small effort. Moreover, we show that our concrete use case can be realized with such an enrichment.

## **Einleitung**

Forschungsdaten (oder Primärdaten) spielen eine wichtige Rolle im Forschungsprozess. Sie fallen bei der Forschung in nahezu allen Bereichen an, bzw. sind der eigentliche Gegenstand der Forschung. Prominente Beispiele sind Umfrageergebnisse, die Ergebnisse von Experimenten, Klimadaten oder große Textkorpora. Auch die eingesetzten Werkzeuge zur Auswertung, erzeugte Zwischenergebnisse oder für die Auswertung aufbereitete Daten zählen zu den Forschungsdaten. Es handelt sich also um einen weit gefassten Begriff, der insbesondere den unterschiedlichen Forschungsdisziplinen Rechnung trägt, deren Bandbreite an verwendeten Forschungsdaten durch die genannten Beispiele skizziert werden soll.

Die Verwaltung und Sicherung der Forschungsdaten obliegt in vielen Fällen den Wissenschaftlern, die sehr unterschiedliche Ansprüche an die Archivierung. Oftmals findet diese sogar eher als Nebeneffekt des Publikationsprozesses und der Kollaboration mit Koautoren statt (Marshall, 2008). Ein Beispiel ist die E-Mail als zentrale Komponente einer Archivierung, wenn E-Mail als Kommunikationsmedium zwischen den Autoren genutzt wird.

Die zuverlässige Sicherung für persönliche Zwecke, aber auch die Möglichkeit, die Daten anderen zugänglich zu machen sind gute Gründe der Bewahrung, Erschließung und Zugänglichkeitmachung von Forschungsdaten. Durch einen strukturierten Zugang zu den Daten kann eine weitergehende Forschung ermöglicht sowie die Transparenz der eigenen Forschungsergebnisse erhalten werden.

Die Thematik, inwieweit eine erhöhte Transparenz und die Möglichkeit zur weiteren Forschung auf den Daten von den Wissenschaftlern tatsächlich gewünscht werden, soll hier nicht weiter vertieft werden. Schon jetzt gibt es entsprechend aufbereitete und zentral archivierte Forschungsdaten, allerdings zumeist mit abweichender Motivation. So fallen in den Naturwissenschaften teilweise Datenmengen an, die von einem einzelnen Wissenschaftler schlicht kaum zu handhaben sind und oft in großen gemeinschaftlichen Projekten erhoben werden, wie zum Beispiel die Ergebnisse von Großexperimenten in der Physik, der Genetik oder der Klimaforschung. Ein anderes Beispiel sind Studien, die oft auch den Charakter einer Publikation haben und die bewusst zur Interpretation und Auswertung anderen Wissenschaftlern zur Verfügung gestellt werden.

### **Verknüpfung von Forschungsdaten und Publikationen**

Die Archivierung von Forschungsdaten ist nur der erste, notwendige Schritt. Für eine effiziente Nutzung der Daten ist es essentiell, dass Verknüpfungen zu den entsprechenden Publikationen bestehen. Analog zur Zitierung anderer Publikationen müsste die Verwendung eines konkreten Forschungsdatensatzes explizit gemacht werden und diese Referenz wiederum den Zugang zu den Forschungsdaten ermöglichen. Wenn solche Referenzen über Systemgrenzen hinweg erfasst und ausgewertet werden können, entstehen effektive Recherchewerkzeuge, ebenfalls analog zu zitierten Publikationen: es ist möglich, auf die zugrunde liegenden Daten zuzugreifen, zur Unterstützung der eigenen Forschung oder zur Verifikation. Zusätzlich können aber auch gezielt weitere Publikationen gefunden werden, die auf den gleichen Daten aufbauen aber eventuell andere Schlüsse ziehen. Schließlich können die Wissenschaftler so mit den Mitteln der Literaturrecherche relevante Datensätze finden, die für die eigene Forschung verwendet werden können. Dies ist gerade in den empirischen Forschungsgebieten wie den Sozialwissenschaften sehr wichtig.

Heute werden Publikationen und Forschungsdaten typischerweise in getrennten Systemen sehr unterschiedlich archiviert. Während sich Systeme im Bibliotheksbereich auf die Publikationen fokussieren, konzentrieren sich Registrierungsagenturen für Forschungsdaten wie *da|ra*<sup>1</sup> auf die Forschungsdaten selbst. Eine Verknüpfung von Forschungsdaten findet sehr selten, generell unvollständig und meist uneinheitlich statt (Boland, et al. 2012).

Im Gegensatz zur relativen Einheitlichkeit der Referenzierung von verwendeter Literatur sehen Referenzen auf Forschungsdaten, so vorhanden, sehr unterschiedlich aus. Manche Autoren referenzieren die Forschungsdaten direkt im Fließtext, während andere Fußnoten oder Überschriften von Tabellen und Grafiken verwenden. Sind Wissenschaftler daran interessiert auf welchen Forschungsdaten Publikationen basieren, ist es somit notwendig Publikationen manuell nach Referenzen zu durchsuchen. Um diesen Prozess zu erleichtern, stellt sich die Universitätsbibliothek Mannheim<sup>2</sup> zusammen mit dem Leibniz Institut für Sozialwissenschaften GESIS<sup>3</sup> und dem Lehrstuhl für Künstliche Intelligenz an der

---

<sup>1</sup> <http://www.gesis.org/dara/>

<sup>2</sup> <http://www.bib.uni-mannheim.de/>

<sup>3</sup> <http://www.gesis.org/>

Universität Mannheim<sup>4</sup> im InFoLiS Projekt<sup>5</sup> unter anderem die Aufgabe Forschungsdaten mit Publikationen zu verknüpfen. Der Fokus in diesem Projekt liegt dabei auf Forschungsdaten und Publikationen aus den Sozialwissenschaften. Um die Extraktion der Referenzen zu automatisieren wurde im Rahmen des Projekts ein Verfahren entwickelt, das Referenzen zu Forschungsdaten in Volltexten extrahiert. Dabei werden automatisch Muster erkannt, die typisch für eine Referenzierung von Forschungsdaten sind (Boland, et al. 2012).

Damit das Einbinden der gewonnen Informationen in die jeweiligen Systeme problemlos ablaufen kann, werden die Verknüpfungen als Linked Open Data veröffentlicht. Dies dient zum einen als einfache gemeinsame Grundlage auf dessen Basis die Systeme selbst die Daten gegebenenfalls transformieren und anschließend laden können. Zum anderen bietet es die Möglichkeit der freien Nachnutzung in weiteren Systemen. Im Folgenden stellen wir vor, wie eine solche Nachnutzung aussehen kann.

### **Integration von Forschungsdaten und Verknüpfungen**

Neben der Verknüpfung von Publikationen und Forschungsdaten ist die anschließende Integration der Forschungsdaten in bibliothekarische Recherchesysteme ein weiteres Ziel im InFoLiS Projekt. Die Umsetzung erfolgt im Resource Discovery System „Primo“<sup>6</sup> der Firma Ex Libris, das an der Universitätsbibliothek Mannheim zum Einsatz kommt. Die hier vorgestellte Vorgehensweise ist aber ohne weiteres verallgemeinerbar und gerade durch den Einsatz von Linked Open Data unabhängig von den im Projekt eingesetzten zugrunde liegenden Systemen.

Bei der Integration der Forschungsdaten in Primo geht es darum, dem Benutzer bei der Detailansicht einer Ressource die Beschreibungen damit verknüpfter Forschungsdaten anzuzeigen und einen Link zu den Forschungsdaten anzubieten. Damit kann ein schneller und einfacher Zugang zu den Forschungsdaten gewährt werden.

---

<sup>4</sup> <http://ki.informatik.uni-mannheim.de/>

<sup>5</sup> <http://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/infolis/>

<sup>6</sup> <http://www.exlibrisgroup.com/category/PrimoOverview/>

Der Link führt zum ursprünglichen Forschungsdatenrepositorium, in dem in gleicher Weise Informationen über verknüpfte Publikationen integriert werden. Dadurch lassen sich die oben skizzierten Anwendungsfälle abbilden.

Für die Integration gibt es grundsätzlich zwei Strategien, die wir im Folgenden als serverseitige und clientseitige Anreicherung bezeichnen.

### **Serverseitige Anreicherung**

Bei einer serverseitigen Anreicherung werden die Beschreibungen der Forschungsdaten und die Verknüpfungen zu den Publikationen direkt in die Datenbank des Recherchesystems geladen. Dazu reicht es im Allgemeinen nicht, die Daten in ein Format zu überführen, das vom Recherchesystem verarbeitet werden kann. Zwar können Beschreibungen der Forschungsdaten analog zu den Beschreibungen der Publikationen geladen und somit die Forschungsdaten zum Bestandteil des Katalogs werden. Für die Verknüpfungen von Publikationen ist aber im Allgemeinen keine Möglichkeit vorhanden. Unter Umständen können Strukturen übernommen oder angepasst werden, die andere Verknüpfungen unterstützen, wie zum Beispiel für die Zusammenführung von Auflagen oder die Verknüpfung von ähnlichen Ressourcen (Recommender-Systeme). Auf jeden Fall sind Anpassungen am Datenschema des Recherchesystems notwendig.

Die Beschreibung der Forschungsdaten weicht durchaus von der Beschreibung einer Publikation ab. Es gibt Gemeinsamkeiten, wie Urheber, Titel und Jahr, aber auch spezifische Informationen, die unter Umständen nicht unter den Tisch fallen sollen, wie einen Erhebungszeitraum.

Ein besonderes Problem der serverseitigen Anreicherung ist damit, dass die erforderlichen Schritte nicht von einem auf ein anderes System übertragen werden können. Je nach Komplexität kann es sogar durchaus sein, dass eine solche Anpassung noch nicht einmal 1:1 auf eine andere Instanz des gleichen Systems übertragen werden kann, da durch die grundsätzlich vorhandenen Anpassungsmöglichkeiten der Systeme einrichtungsspezifische Unterschiede bestehen können. Ein weiterer Nachteil kann auch die notwendige Datenreplikation sein, insbesondere die daraus resultierende Frage nach der Aktualität.



Ein entsprechender Updatemechanismus ist deshalb vorzusehen. Grundlegende Änderungen zum Beispiel beim Format oder den anzuzeigenden Daten erzeugt einen relativ hohen Wartungsaufwand. Nicht zuletzt müssen ggf. auch Lizenzvereinbarungen beachtet werden, wenn externe Daten auf diese Weise dupliziert werden.

Auf der anderen Seite ist es vorteilhaft, dass die externen Informationen direkt in die Datenbank geladen werden. So verhalten sie sich wie jegliche andere Art von Daten im System und können dadurch auch recherchierbar gemacht werden. Auch andere Funktionen wie die Facettierung können genutzt werden, um zum Beispiel gezielt nach Publikationen mit Forschungsdaten zu suchen.<sup>7</sup> Eine serverseitige Anreicherung ist generell erforderlich, wenn die Suche innerhalb des Recherchesystems verbessert werden soll, zum Beispiel durch die Integration zusätzlicher Erschließungsdaten.

### **Clientseitige Anreicherung**

Unter einer clientseitigen Anreicherung verstehen wir, dass die zusätzlichen Daten erst zur Laufzeit im Client eingebunden werden, im Fall von Primo also zum Beispiel per JavaScript im Browser des Benutzers. Dementsprechend ist eine Transformation der Daten in das interne Format des Recherchesystems nicht nötig.

Natürlich bedeutet eine Anreicherung der Daten in der Präsentationsschicht, dass die zusätzlichen Daten nicht für die Recherche genutzt werden können. Trotzdem lassen sich wichtige Anwendungsfälle unterstützen.

Die clientseitige Anreicherung bietet den Vorteil, dass zusätzliche Informationen auf einfache Weise zur Verfügung gestellt werden können, auch ohne direkte Unterstützung des Recherchesystems. Bei geschickter Implementierung kann die systemabhängige Einbindung in den Client von der eigentlichen Umsetzung der Anreicherung getrennt werden, so dass eine weitgehende Systemunabhängigkeit erreicht wird. Durch die direkte Einbindung der externen Datenquellen stehen auch stets die aktuellsten Daten zur Verfügung.

---

<sup>7</sup> Das lässt sich allerdings auch dadurch erreichen, dass lediglich die Information, ob Forschungsdaten zur Publikation vorliegen, im System vermerkt wird.

Um dem Problem zu entgehen, dass der Datenprovider zum Flaschenhals wird, müssen allerdings in der Praxis im Allgemeinen Proxy-Caches eingesetzt werden. Für die Frage, ob eine client- oder serverseitige Anreicherung verwendet wird, kann auch die gewünschte Präsentation für den Benutzer eine Rolle spielen. Viele clientseitige Anreicherungen verzögern die Anzeige. Zwar können die Daten asynchron im Hintergrund geladen werden, doch die Verzögerung, mit der die zusätzlichen Daten dann auf der Seite eingeblendet werden, ist für den Benutzer deutlich wahrnehmbar und unter Umständen nicht gewünscht.

Ähnlich wie bei der serverseitigen Anreicherung gibt es auch für diesen Ansatz typische Anwendungsfälle. Ein Beispiel ist das Einbinden von Wikipedia-Artikeln zu Publikationen oder Autoren. Die Informationen werden normalerweise nicht für die Recherche verwendet, sondern lediglich dem Benutzer bei Bedarf zur Verfügung gestellt, um den Komfort zu erhöhen. Eine Generalisierung unseres Anwendungsfalls der Integration von Forschungsdaten ist generell das Einbinden von Empfehlungen, die sich auf eine gerade angezeigte Publikation beziehen. Derartige Informationen unterstützen gerade nicht die Suche nach Publikationen, sondern das Browsen im Bestand.

Bei welchem Anwendungsfall welche Art der Anreicherung am besten verwendet wird, hängt demnach von einigen Faktoren ab. Diese sollten von der jeweiligen Institution, die eine Anreicherung plant, im Voraus betrachtet und untersucht werden.

### **Linked Data als Infrastruktur**

Unabhängig von der Art der Anreicherung wäre es wünschenswert, die externen Datenquellen möglichst einheitlich einbinden zu können. Bei der clientseitigen Anbindung fällt das noch mehr ins Gewicht, da dort direkt mit den Daten gearbeitet wird und jede Unterstützung eines spezifischen Formats zusätzlichen Aufwand bedeutet. Ein weiteres Desideratum wäre eine einheitliche Identifizierung der durch die Daten beschriebenen Ressourcen, um die Integration zu vereinfachen und Probleme beim Zusammenführen der Daten durch uneinheitliche Identifizierung zu vermeiden.

Linked Data ist das Konzept einer derartigen homogenen Datenlandschaft, die den einfachen Austausch und die Integration verschiedenster Daten unterstützt. Dazu werden standardisierte Techniken eingesetzt, die direkt auf der Infrastruktur des World Wide Web aufsetzen. Bei Linked Data, gibt es vier einfache Grundprinzipien (Bizer, Heath und Berners-Lee 2009):

1. Weltweit eindeutige Identifier (URIs) benutzen um Ressourcen zu identifizieren.
2. Mittels HTTP-URIs können Ressourcen referenziert und nachgeschlagen werden.
3. Daten über eine Ressource werden über dessen URI einheitlich bereitgestellt.
4. Durch Bezüge (Links) zu anderen Daten entsteht ein Datennetz, so dass weitere Daten gefunden werden können. Auf diese Weise können einmal integrierte Daten unmittelbar zur Nachnutzung durch andere zur Verfügung gestellt werden.

Was bedeuten diese Prinzipien für die Integration von Forschungsdaten und Publikationen? Zuerst müssen Publikationen sowie Forschungsdaten mit HTTP-URIs versehen werden. Dies ist wird zwar üblicherweise von aktuellen Systemen nicht direkt unterstützt, allerdings gibt es im Allgemeinen lokale Identifier, die durch einen URIPräfix zur HTTP-URI erweitert werden können, so dass zum Beispiel URIs der folgenden Form entstehen: [http://link.bib.uni-mannheim.de/primo/MAN\\_ALEPH001437414](http://link.bib.uni-mannheim.de/primo/MAN_ALEPH001437414).

MAN\_ALEPH001437414 ist dabei die interne Record-ID des Datensatzes in Primo. Durch die Kontrolle der Domain bib.uni-mannheim.de ist sichergestellt, dass die URI weltweit eindeutig bleibt. Wichtig ist dabei allerdings, dass die internen Identifier auch stabil bleiben. Die Daten **sollten** dauerhaft zur Verfügung stehen (das kann zum Beispiel in Primo nicht garantiert werden, da Daten von Drittanbietern eingebunden werden und zum Beispiel durch Abbestellungen oder Aussonderungen Einträge aus dem Katalog verschwinden) und ein Identifier **darf nicht** zu einem späteren Zeitpunkt für eine andere Ressource verwendet werden, selbst wenn die ursprüngliche Ressource nicht mehr vorhanden ist (das ist hoffentlich bei allen Daten in Primo der Fall, aber bei den Drittanbietern kaum endgültig zu klären). Mit der Vergabe von solchen HTTP-URIs sind somit die ersten beiden Linked Data Prinzipien erfüllt.

Um Informationen standardisiert bereitzustellen, werden die Publikations- und Forschungsdaten nach RDF (Resource Description Framework) transformiert.

RDF ist ein sogenanntes Metamodell, mit dem beliebige Datenmodelle (Ontologien) und die zugehörigen Daten einheitlich repräsentiert werden können. Für Publikationen und Forschungsdaten stehen bereits verschiedene, mehr oder weniger mächtige, Ontologien zur Verfügung. Für grundlegende Angaben zu den Ressourcen kann zum Beispiel Dublin Core<sup>8</sup> verwendet werden. Speziellere Vokabularien sind zum Beispiel DDI<sup>9</sup> für Forschungsdaten in den Sozialwissenschaften. Die besondere Stärke von RDF liegt unter anderem darin, dass diese Ontologien für eigene Zwecke erweitert und verfeinert werden können, ohne dabei die Interoperabilität der Daten zu reduzieren.

Sobald die Publikations- und Forschungsdaten in RDF vorliegen, können zu einer Ressource gehörende Daten direkt unter der entsprechenden URI zur Verfügung gestellt werden. Idealerweise wird über eine sogenannte Content-Negotiation erkannt, wenn ein Mensch mit einem normalen Webbrowser auf die URI zugreift und in diesem Fall direkt auf die menschenlesbare Repräsentation der Daten, zum Beispiel in Primo, weitergeleitet.

Wenn die Daten noch Verweise zu anderen Daten enthalten, zum Beispiel zu Forschungsdaten, die einer Publikation zugrunde liegen, dann sind alle vier Prinzipien erfüllt.

Als Teilprojektergebnis werden im InFoLiS Projekt die Verknüpfungen von Publikationen zu Forschungsdaten außerhalb der Systeme als Linked Open Data zur Verfügung gestellt, so dass sie unabhängig von den Ursprungsdaten beliebig nachgenutzt werden können. Über eine Suchschnittstelle (SPARQL) können die Daten gezielt abgefragt werden.

## **Ablauf**

Für die Integration von Publikationen und Forschungsdaten im Projekt InFoLiS kommt die clientseitige Anreicherung zum Einsatz. Abbildung 1 illustriert den Ablauf einer typischen Benutzeranfrage und die Schritte zur Anreicherung. Zunächst kommt der Benutzer auf die Detailansicht einer Publikation in Primo, entweder als Ergebnis einer vorherigen Suche, oder durch den Aufruf der identifizierenden URI, die

---

<sup>8</sup> <http://dublincore.org/>

<sup>9</sup> <http://www.ddialliance.org/>

menschliche Benutzer auf die entsprechende Primo-Seite verweist. Die Seite enthält JavaScript-Code, der die eigentliche Anreicherung vornimmt.

Dazu muss zunächst festgestellt werden, welche Publikation auf der aktuellen Seite angezeigt wird. Diese Daten können direkt der Seite entnommen werden, besonders einfach geht das, wenn sie bereits strukturiert vorliegen, zum Beispiel über eingebettetes RDF (RDFa). Ist die URI der gerade angezeigten Publikation bekannt, dann kann über eine Suchanfrage in SPARQL nach Links zu Forschungsdaten gefragt werden. Hier wird also die erste externe Datenquelle eingebunden.

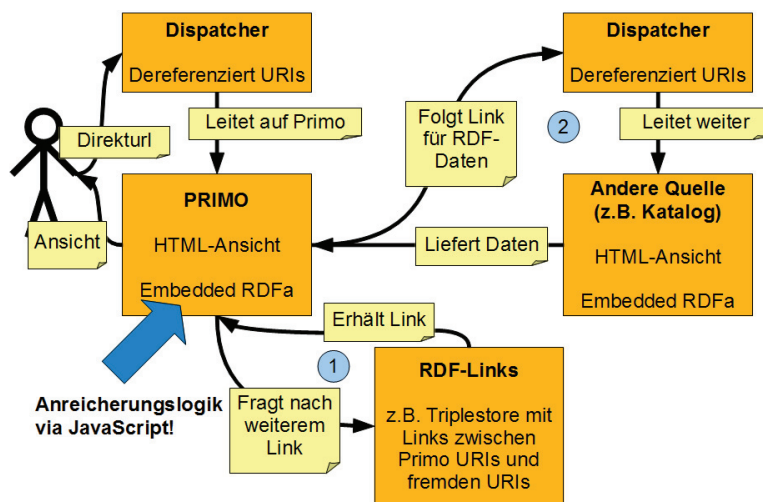


Abbildung 1: Ablauf einer clientseitigen Anreicherung in Primo

Die Links, also hier speziell die URIs von Forschungsdaten, können direkt dereferenziert werden und führen zu einer weiteren externen Datenquelle. Dort werden wiederum die gewünschten Informationen, die zur Anzeige der Forschungsdaten benötigt werden, zur Verfügung gestellt. In diesem Beispiel wird dabei sogar auf eine Unterscheidung der Anfragen verzichtet und einfach immer eine Webseite mit eingebettetem RDFa verwiesen.

Im letzten Schritt werden die erhaltenen Daten dynamisch in die Detailseite eingebunden und für den Benutzer erscheinen die gewünschten Links zu Forschungsdaten.

## **Fazit**

Linked Data bietet eine ideale Grundlage zur Integration heterogener Daten und sichert insbesondere die leichte Nachnutzbarkeit auch außerhalb des ursprünglich anvisierten Zwecks. Die Bereitstellung der Daten und die Integration selbst kann dabei immer noch auf verschiedene Arten erfolgen, mit entsprechenden Vor- und Nachteilen. Wir haben nach server- und clientseitiger Datenanreicherung unterschieden und gezeigt, dass die clientseitige Anreicherung weitgehend systemunabhängig und mit minimalem Eingriff in die bestehenden Systeme erfolgen kann. Für bestimmte Szenarien ist diese Art geradezu prädestiniert, insbesondere zur Unterstützung des Browsings der Benutzer durch den Datenbestand, im Gegensatz zur gezielten Suche. Erfolgt die Bereitstellung der Daten direkt in den Webseiten des Ursprungssystems, ist nicht einmal eine Duplizierung der Daten notwendig. Natürlich können bestimmte Anwendungsfälle, insbesondere die Suche in den Daten, dann nicht unterstützt werden, aber die Vorteile überwiegen in vielen Fällen deutlich. Wir bezeichnen eine solche Art der Umsetzung als „Pareto-Implementierung“. Auch wenn die serverseitige Anreicherung und auch die Bereitstellung der Daten in dedizierten Systemen (idealerweise sogar mit SPARQL-Schnittstelle) unbestreitbare Vorteile bietet, schreckt doch gerade der erhebliche Aufwand und die (vorerst) notwendige Datenduplizierung viele Einrichtungen vom konsequenten Einsatz von Linked Data ab.

Ein System das Linked Data bereits erfolgreich integriert ist lobid.org. Dabei werden Publikationsdaten als Linked Data veröffentlicht und unter Benutzung weiterer Linked Data Quellen angereichert (Jansen und Christoph 2012). Wir hoffen, dass wir zeigen konnten, dass diese Bedenken für die Pareto-Implementierung in den meisten Fällen unbegründet sind.

## Literaturverzeichnis

- Bizer, Christian, Tom Heath, und Tim Berners-Lee. „Linked Data - The Story So Far.“ *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009: 1-22.
- Boland, Katarina, Dominique Ritze, Kai Eckert, und Brigitte Mathiak. „Identifying references to datasets in publications.“ *Proceedings of TPDL 2012: Theory and Practice of Digital Libraries*. 2012.
- Jansen, Heiko, und Pascal Christoph. „Dynamische Kataloganreicherung auf Basis von Linked Open Data.“ 101. *Deutscher Bibliothekartag* . Hamburg, 2012.
- Marshall, Catherine. "From writing and analysis to the repository: taking the scholars' perspective on scholarly archiving." *Proceedings of the 8<sup>th</sup> ACM/IEEE-CS joint conference on Digital libraries (JCDL' 08)*. ACM, New York, NY, USA, 2008. 251-260.
- Pfeffer, Magnus. „Automatische Vergabe von RVK-Notationen mittels fallbasiertem Schließen.“ 97. *Deutscher Bibliothekartag in Mannheim 2008 - Wissen bewegen. Bibliotheken in der Informationsgesellschaft*. Frankfurt: Klostermann, 2009. 245-254.
- Ritze, Dominique, und Kai Eckert. „Data Enrichment in Discovery Systems using Linked Data .“ *Proceedings of the 36th Annual Conference of the German Classification Society on Data Analysis, Machine Learning and Knowledge Discovery* . 2012, to be published.

## **Semantische Digitale Bibliotheken Bibliotheken als Mehrwertproduzenten in vernetzten Wissensarchitekturen**

Stefanie Sorge

Hochschule Biberach, Bibliothek

### **Zusammenfassung**

Das soziale Zusammenleben sichert das menschliche Miteinander. In jeder Gesellschaftsform wurde das Zusammenleben über bestimmte Sozial- und Kulturtechniken geregelt. Auch wer Zugang zu Informationen hatte und wer nicht, wurde z. B. über die Kulturtechnik Lesen durchgesetzt. Die dezentralen Netzwerke der Computergesellschaft integrieren nun immer mehr Menschen. Demzufolge fließen auch immer mehr Informationen. Während früher Informationen künstlich verknüpft wurden, stehen heute Informationen in unüberschaubaren Mengen zur Verfügung. Sind bei dem einen Anbieter bestimmte Informationen knapp, ist die Wahrscheinlichkeit hoch, bei einem anderen Anbieter die gesuchte Information zu finden. Neben den tradierten Gedächtniseinrichtungen gibt es eine Vielzahl von neuen Akteuren auf den Wissens- und Informationsmärkten. Kollektiv, aber nicht immer durch kollaborative Zusammenarbeit, wird versucht, die Informationsmassen zu ordnen. Unterschiedliche Informationsarchitekturen greifen teilweise ineinander und schaffen neuartige hybride Strukturen.

Bibliotheken sind Akteure in diesen Netzwerken und haben das Potenzial, Qualitätsinstanzen für das zukünftige semantische Web zu werden. Für dieses Ziel sind mehrere Aspekte zu beachten, insbesondere Prozesse der Ordnung und Prozesse der Lernenden Organisation.

*Digitale Bibliothek, Management, Systemtheorie, Internet, Vernetzung, Gesellschaft, Wissensmanagement*

### **Abstract**

Social life is depended upon interpersonal communication. All social systems have relied on particular social and civilization techniques. For example, the civilization technique of reading determined who had access to information.

Today increasing numbers of people are able to gain access to computer networks and an increasing amount of information is being produced. While information was kept scarce in former times, an incalculable amount of information is available today. If a particular piece of information cannot be provided by a special institution, most likely it can be obtained through some other source. Apart from the traditional libraries, public domain databases and the like a larger number of specialized providers are active in the knowledge and information markets.



Jointly, but not necessarily in co-operation the masses of information are classified. Information architectures are partially integrated and new hybrid structures arise. Libraries also participate in these networks and could become higher authorities controlling the quality of semantic webs. Several aspects should be noted for this purpose, in particular processes of order and processes of learning organizations.

*digital library, management, systems theory, internet, networking, society, knowledge management*

## **1 Einführung/Forschungsstand**

Eine Möglichkeit die Entwicklung von Gesellschaften zu verstehen, ist die Einteilung in Epochen: Tribale Gesellschaft, Antike Gesellschaft, Moderne Gesellschaft, Computergesellschaft. Jeder Epoche kann ein zentrales Leitmedium zugeordnet werden. Soziologen, wie auch Medienwissenschaftler, versuchen über den Einfluss von Medien, die Wandlungsprozesse in Gesellschaften zu erklären. Wir leben heute in der Computergesellschaft. Diese zeichnet sich durch eine hohe Dynamik, Gleichzeitigkeit und Auflösung von Grenzen aus. Zugleich werden neue Grenzen sichtbar: Grenzen des Wachstums, Grenzen des Vertrauens, Grenzen der Globalisierung, Grenzen der Bildung, Grenzen des Wissens, Grenzen der Medizin, Grenzen der Wahrnehmung. Als Organisationsprinzip kann die Vernetzung angeführt werden. Die Computergesellschaft ist komplex, denn bei Verunsicherung greifen die Gesellschaftsmitglieder auf gelernte Kultur- und Sozialtechniken vergangener Epochen zurück (z. B. Aberglaube, Esoterik, Starkult, Applekult, Heldenverehrung, Krieg, Vorurteil, Kritik, Verleugnung, Ablehnung, Reflexe, Vertrauen, Hörigkeit).<sup>1</sup>

Bibliotheken sind institutionalisierte Informationsarchitekturen, welche Informationen sammeln, erschließen und durch Kontextualisierung, Aggregation, Ordnung, Organisation zu Wissensarchitekturen transformieren. Als Informationen werden in Anlehnung an das DGI-Positionspapier, Daten in bestimmten Mustern bezeichnet. Wissen besteht aus Informationen im Kontext.<sup>2</sup> Durch standardisierte Verfahren kann dieser Kontext durch Maschinen syntaktisch verarbeitet werden. Menschen sind den Maschinen aus mehreren Gründen überlegen: können Mehrdeutigkeiten erfassen, Selbstorganisation des Denkens, Assoziationsfähigkeit, stabile selbsterhaltende freibewegliche Organisationsstruktur Körper und geteilte Wissensbasis über die Welt (z. B. Ein Stein wirkt wie ein fester Körper).

---

<sup>1</sup> Vgl. Baecker, Dirk, Sternstunde Philosophie, vgl. McLuhan, Marshall (1964), Understanding media.

<sup>2</sup> Vgl. DGI e. V. (2010), Positionspapier, 3.

Menschen greifen auf ihre Wissensbasis intuitiv zu. Einer Maschine muss das strukturierte Erfassen von Informationen mühsam beigebracht werden.

Semantische Digitale Bibliotheken können gespeichertes Wissen repräsentieren. Jedoch simulieren die syntaktisch arbeitenden Suchinterfaces menschliches Verstehen. Syntaktische Technologien bilden Informationsangebote über formalisierte mathematisch beschreibbare Muster ab. Semantische Digitale Bibliotheken beschreiben oftmals Applikationen (z. B. Expertensysteme in der Medizin), welche auf syntaktischen Technologien basieren, aber nicht dem menschlichen Verstehen, nach heutigem Stand der Technik, gleichzusetzen sind.

Neben den tradierten Informationsarchitekturen der Bibliotheken, existierten immer auch informelle Beziehungs- bzw. Informationsnetzwerke (z. B. Forschergruppen, Freunde, Familie, Firmenmitarbeiter). Viele Informationsarchitekturen im Internet zeigen die alten Organisationsstrukturen, jedoch entstehen auch neuartige hybride Informationsarchitekturen z. B. Social Web. Um hybride dezentral organisierte Computernetze zu verstehen, müssen die Prozesse sozialer hybrider Kollektive untersucht werden. Wieso kann ein Netzwerk effizienter sein, obwohl sich die Teilnehmer nicht kennen? Welche Regeln liegen diesen Strukturen zugrunde? Wie lernen künstliche neuronale Netze? Was bedeutet Mehrwert in den neuen Netzkulturen? Welche neuen Sozialtechniken entstehen? Wie können tradierte Informationsarchitekturen Anschluss an die soziale Realität herstellen? Welche neuen Services- und Dienstleistungen benötigen die Computergesellschaften? Welche Rolle spielt der Mensch in künstlichen neuronalen Netzwerken und wie können Semantische Digitale Bibliotheken als offenen, dezentrale Wissensarchitekturen konzipiert werden?

## **2 Darstellung der Methodik**

Zur Analyse des Untersuchungsgegenstandes, Semantische Digitale Bibliotheken und Mehrwertpotenziale vernetzter Wissensarchitekturen, wurde die systemtheoretische Analyse gewählt. Hierfür wurde systemtheoretische Literatur aus den Bereichen Soziologie, Wirtschaft und Psychologie analysiert. Weiterhin wurde Literatur aus den Medien- und Informationswissenschaften genutzt.

In der operationalen funktionalen Systemtheorie nach *Luhmann* gibt es eine Beschreibungslücke. Nach *Luhmann* besteht die Gesellschaft aus funktionalen Systemen, welche füreinander Umwelt sind. Menschen kommen bei *Luhmann* nicht vor.

Menschen sind zusammengesetzte Systeme aus biologischen und psychischen Systemen und operieren über Leben und Bewusstsein. Kritisch ist zu hinterfragen, wenn es keine sinnlich, kreativen Individuen gibt, wie sollen sich soziale Systeme überhaupt bilden können? Wie kann es sein, dass Sinn und Vertrauen einfach vorausgesetzt werden? *Luhmanns* Beschreibung der Gesellschaft kann nicht die Prozesse innerhalb und zwischen den unterschiedlichen Systemarten erklären (z. B. Informationsarchitektur physikalisches Gehirn und psychische virtuelle Informationsarchitektur Denken/Geist). Weiterhin seien nach *Luhmann*, soziale Systeme autopoietischen Systemen.<sup>3</sup> Dem widerspricht *Hejl* und kritisiert ebenfalls das Fehlen des Individuums in *Luhmanns* Theorie. Nach *Hejl* organisieren sich soziale Systeme nicht selbst, da sie sich nicht in operativ geschlossener Weise erzeugen. Soziale Systeme sind nicht selbsterhaltend, da sie an die Lebensdauer ihrer Komponenten gebunden sind. Für *Hejl* bestehen soziale Systeme aus Systemmitgliedern und aus einer Systemorganisation.<sup>4</sup> Teilnehmer des Systems z. B. Bibliotheksteam, werden die Komponenten nur, wenn sie die gleichen Regeln teilen und auch tatsächlich nach diesen Regeln handeln (z. B. Arbeitsvertrag unterschreiben, Akzeptanz hierarchischer Organisation, Katalogisierungsregeln, Unternehmenskultur, Einsatz IT-Systeme ect.). Daher greift die Führungstheorie, welche eine Organisation als mechanische Maschine betrachtet zu kurz. Es sind eben nicht die Komponenten beliebig ersetzbar.

*Maturana* und *Varela* beschreiben in ihrer Theorie der autopoietischen Systeme lebende Systeme als selbsterhaltend und selbstregulierend. Weiterhin operieren lebende Systeme informell, funktional geschlossen, jedoch energetisch offen.<sup>5</sup> Durch *Ciompis* Theorie der fraktalen Affektlogik kann erklärt werden, wie an den Systemgrenzen zwischen biologischen und psychischen Systemen aus dem Denken Aufmerksamkeit und tatsächliches Handeln erzeugt werden. Gedanken sind mit Emotionen verbunden und erzeugen strukturelle Kopplungen innerhalb

---

<sup>3</sup> Vgl. *Luhmann, Niklas* (1984), Soziale Systeme, 168.

<sup>4</sup> Vgl. *Hejl, Peter M.* (1987), Konstruktion der sozialen Konstruktion, 320.

<sup>5</sup> Vgl. *Maturana, Humberto R./Varela, Francisco J.* (1980), Autopoiesis and cognition, 118-123.

unterschiedlicher Systemarten. Durch diese energetische Offenheit können voneinander unabhängige Individuen auf bestimmte Handlungsmodi gelenkt werden. Dies geschieht unter Mitvollzug des erleidenden Systems. So erschafft sich jeder Mensch durch seine eigenen Gedanken und durch eigene Emotionen seine subjektive Wirklichkeit. Seine eigene Wissensarchitektur.

Daher sind nach *Ciampi* Emotionen Mittler zwischen den Systemgrenzen. Denn Emotionen wirken auf die Fühl-, Denk- und Verhaltensprogramme.<sup>6</sup>

Dies ist die Basis für kreative Austauschprozesse zwischen den Menschen als soziale, emotionale, denkende, fühlende, kreative, intuitiv verstehende, mit Bewusstsein ausgestattete und zu moralischem, ethischem, mitfühlendem Verhalten fähigen Wesen. Der Mensch ist damit selbst eine der stabilsten Informationsarchitekturen, seit einigen Millionen Jahren. Die soziale Realität, begründet sich auf geteilter Wirklichkeit. Die soziale Wirklichkeit basiert auf geteilten Wissensarchitekturen.<sup>7</sup>

In den Computernetzwerken ist die Frage des Akteurs und des Sozialen wichtig. Wer handelt eigentlich? Was verbindet die Akteure untereinander? *Latour* untersucht in seiner Akteur-Netzwerk-Theorie, wie Informationen zu Wissen transformiert werden. Dabei zeigt er auf, dass Menschen in Netzwerken handeln und Kollektive mit Nicht-Menschen bilden und sogar zu Hybriden verschmelzen. *Latour* schlägt daher vor, den Begriff der Gesellschaft durch den Begriff der Kollektive zu ersetzen.<sup>8</sup>

*Latours* Akteur-Netzwerk-Theorie erklärt, wie passive Systeme (z. B. Apparaturen) auf Grundlage des aktiven Systems Mensch in Netzwerken handeln können. Desweiteren zeigt *Latour* auf, wie durch die Nicht-Lebewesen erst neue Informationen entstehen können. *Latour* zeigt wie neue hybride Informationsarchitekturen (z. B. Labore), entstehen und diese wiederum unsere Wissensbasis bereichern. Auch *Hejl* verweist darauf, wie die passiven Systeme (z. B. Hierarchien) auf Basis der aktiven Systeme (z. B. Bibliotheksmitarbeiter) eine Informationsarchitektur bilden.<sup>9</sup>

---

<sup>6</sup> Vgl. *Ciampi, Luc* (2005), Die emotionalen Grundlagen des Denkens, 68.

<sup>7</sup> Vgl. *Sorge, Stefanie* (2012), Angst und Leistung in Teams der Medienwirtschaft, 98.

<sup>8</sup> Vgl. *Latour, Bruno* (2002), Die Hoffnung der Pandora, 374 f.

<sup>9</sup> Vgl. *Latour, Bruno* (2006), Über technische Vermittlung, 496, vgl. *Hejl, Peter M.* (1994), Die Entwicklung der Organisation von Sozialsystemen und ihr Beitrag zum Systemverhalten, 113-114, vgl. *Sorge, Stefanie* (2011), Semantische Digitale Bibliotheken, 20-22.

### 3 Ergebnisse der eigenen Arbeit

Mithilfe der Systemtheorie können zunächst die komplexen Verflechtungen zwischen unterschiedlichen Systemarten erklärt werden. In vernetzten Computersystemen handeln Menschen, Maschinen und Hybride zunehmend über kollektive hybride Vernetzungen miteinander. Es ist wichtig zu erkennen, dass der Mensch als kreative Inputvariable eine besondere Rolle in den hybriden Kollektiven einnehmen kann. Handeln geht jedoch in dezentral technosozialen Netzwerken auch von Maschinen und Hybriden aus. Diese verbinden sich zu immer wieder neuen spontan entstehenden Systemkomponenten. Daneben können die einzelnen Komponenten auch längerfristig Mitglied einer bestimmten Informationsarchitektur sein, wie z. B. ein Hybrid aus Bibliotheksmitarbeiter und aDIS BMS (Bibliotheksmanagementsystem). Das Ergebnis dieser Überlegungen kann in diesem Beitrag in der *Abbildung: Semantische Digitale Bibliotheken - Bibliotheken als Mehrwertproduzenten in vernetzten Wissensarchitekturen* - nachvollzogen werden.

Wie Semantische Digitale Bibliotheken in vernetzten Wissensarchitekturen Mehrwert produzieren können, ist eine schwierige Frage. Das Internet kann als ein Netzwerk aus Informationsarchitekturen begriffen werden. Jede Architektur besitzt eine eigene Infrastruktur und Organisation. In der Grafik sind diese Netzwerke als baumartige Verästelungen dargestellt. Jede Astgabelung bildet einen Knoten. Jeder Knoten ist eine Informationsarchitektur. Je mehr Informationen organisiert werden müssen, desto höher ist die Wahrscheinlichkeit, dass die Informationsarchitekturen durch ihre Komponenten angepasst werden. So können auf jedem Ast weitere unterschiedliche Informationsarchitekturen an den Knotenpunkten der Astgabeln sitzen und entstehen. Jeder Zweig stellt vereinfacht eine Information dar. Diese Information findet sich in den darüber liegenden Kreisen der Wissensarchitekturen wieder. In der Art wie sich die Informationsarchitekturen ändern, entwickeln sich auch die Wissensarchitekturen.

Teilnehmer an Informationsarchitekturen wie z. B. Südwestdeutscher Bibliotheksverbund können Lebende, Nichtlebende oder Hybride nur werden, wenn sie gleiche Wirklichkeitskonstruktionen (z. B. Gemeinsame Normdatei, Umstellung auf neues Bibliotheksmanagementsystem aDIS BMS, standardisierte technische Schnittstellen) ausgebildet haben und die einzelnen Akteure auch tatsächlich auf Grundlage dieser Organisation handeln.

Der Kreis der Wissensarchitekturen ist offen dargestellt, da sich die Informationsmengen ständig vergrößern und damit auch das potenzielle Wissen. So stieg allein der Anteil der Haushalte mit Zugang zum Internet weltweit im Jahr 2009 auf 27,2 %<sup>10</sup>. In Deutschland hatten im Jahr 2011 bereits 83 % der Haushalte Zugang zum Internet<sup>11</sup>. Der Umsatz von E-Books in den USA betrug 2009 0,551 Mio. USD und wird laut Prognose von *PwC* im Jahr 2015 auf 5,571 Mio. US Dollar steigen<sup>12</sup>.

---

<sup>10</sup> *Itu.int*, in: statista.com, 08.08.2012

<sup>11</sup> *Eurostat*, in: statista.com, 08.08.2012

<sup>12</sup> *PwC*, in: statista.com, 08.08.2012



Quelle: Eigene Darstellung

In Deutschland fiel die Buchproduktion im Printbereich von 84.351 Stück in 2010 auf 82.048 Stück in 2011.<sup>13</sup> Weitere Indikatoren für die Zunahme von Informationen sind die Absatzzahlen für Smartphones. Weltweit wurden in 2010 schon 296,95 Mio. Stück und im Jahr auf 2011 bereits 472 Mio. Stück verkauft<sup>14</sup>. Ebenfalls stiegen allein die Domainzahlen mit der Endung DE in Deutschland in 2011 auf 14.819.132.<sup>15</sup> Auch die steigenden Erstabsolventenquoten legen den Schluss nahe, dass mehr wissenschaftliche Informationen in Zukunft produziert werden.<sup>16</sup>

Die Ordnungstiefen der Wissensarchitekturen soll hier schematisch aus vier Ebenen bestehen. Die Strukturen können in der Realität auch vielfältiger sein. Die Informationsarchitekturen sind die Basis für die Wissensarchitekturen. Da die Informationsflut für die Akteure erst durch eine qualitative Aufbereitung gut nutzbar wird, ist die Gesellschaft auf Qualitätsinstanzen angewiesen. Diese unterschiedlichen Informationsarchitekturen bestehen aus Infrastrukturen und verschiedenartigen Komponenten.

In der Grafik wurden fünf zentrale Informationsarchitekturen abgebildet und weiterhin zwei Hybridstrukturen. Daneben können weitere Strukturen existieren.

#### **Informationsarchitektur: Linear Mono, zentral und geschlossen**

Linear Mono bezeichnet Qualitätsinstanzen durch Verknappung von Informationen. Bibliothekskataloge haben nur Literatur gesammelt, welche professionell gesicherte Filterinstanzen des Wissenschaftsbetriebs durchlaufen hatten. Die Information absolviert einen einförmigen Weg von einer Haltestelle zur Nächsten. Auf diesem Weg durchläuft die Information Knotenpunkte und wird mit Zusatzinformationen durch professionelle Qualitätsinstanzen angereichert. Die Organisationsstruktur ist zentral und geschlossen. Die Benutzer haben von außen keinen Einfluss auf die Bereitstellung der Informationen und auch nicht auf die Entwicklung von Applikationen.

---

<sup>13</sup> Börsenverein Deutscher Buchhandel, in: statista.com, 08.08.2012

<sup>14</sup> Gartner.com, in: statista.com, 08.08.2012

<sup>15</sup> Denic, in: statista.com, 08.08.2012

<sup>16</sup> Statistisches Bundesamt, in: statista, 08.08.2012



### **Informationsarchitektur: Linear Multi, dezentral und geschlossen**

Linear Multi kann die tradierten Qualitätsinstanzen der Bibliotheksaggregatoren wie z. B. Südwestdeutscher Bibliotheksverbund, Verlage oder Datenbanken beschreiben. Auch hier gibt es einen professionellen Berufsstand, welcher durch Informationsverknappung die Qualität erhöht. Es werden Indexe hinzugekauft oder selber erstellt.

Es kommt zu einer Ballung von bestimmten Inhalten und Themenfeldern. Für den Benutzer sind die Angebote frei, durch Mitgliedschaft oder durch Einkauf in die Plattform zugänglich. Die Benutzer dieser Systeme haben weder auf die Erschließung der Information noch auf die Ausgestaltung der Applikationen direkten Einfluss.

### **Informationsarchitektur: Wirbel Mono, dezentral und geschlossen**

Wirbel Mono verweist auf Aggregatoren wie z. B. Google, die Informationen wie Linear Multi aus unterschiedlichen Quellen zusammenführen. Da die zur Verfügung stehenden freien Informationsquellen im Web aggregiert werden, kann die Infrastruktur als Wirbel bezeichnet werden, der sich aus den dezentralen Quellen zu einem Wirbel bündelt. Die Beteiligung der Benutzer ist indirekt möglich durch Auswertung der freiwillig bereitgestellten Benutzerinformationen.

### **Informationsarchitektur: Universal Multi, dezentral und offen**

Universal Multi zeigt eine Informationsinfrastruktur gleich dem Semantic Web, welcher sich durch einen freien Informationsaustausch auszeichnet wie z. B. Anbieter in der Linked Open Data Cloud. Die Informationen sind mit unterschiedlichen Informationsquellen vernetzt. Jeder hat Zugang zu den Daten und es können neuartige Applikationen, auch durch branchenferne Akteuren, entwickelt werden.

### **Informationsinfrastruktur: Wirbel Multi, dezentral und offen**

Wirbel Multi zeigt eine Informationsinfrastruktur gleich dem Social Web z. B. Twitter. Die Wirbelmetapher steht für das Thema Antrieb. Eine Art Kreativschmiede für Informationen. Hier zirkulieren chaotische Strukturen und verbinden sich in dynamischen, natürlichen Austauschprozessen zu potenziellen neuen innovativen Ideen. Auch diese Ideen werden als Information im äußeren Wissensring abgelegt.

Die Wahrscheinlichkeit, dass diese Informationen durch die tradierten Qualitätsinstanzen gefunden werden, ist gering, solange diese Informationen nicht über bessere Wissensorganisationen erschlossen werden.

#### **Hybrid-Informationsinfrastruktur: Dezentral Multi, dezentral und offen**

Dezentral Multi oder die Soziale Semantische Digitale Bibliothek z. B. DBpedia mit der vorgelagerten Wikipedia ist eine Hybridstruktur zwischen Linear Mono Qualitätsinstanzen wie z. B. Kataloge und dem Social Web wie z. B. Wikipedia. Die Informationen werden auch intellektuell erschlossen. Der soziale Aspekt des Informationsaustauschs steht im Vordergrund. Die Usability der Systeme und auch die Qualitätsaufbereitung der Informationen werden durch die Benutzer aktiv mitgestaltet.

#### **Hybrid-Informationsinfrastruktur: Zentral Multi, zentral und geschlossen**

Zentral Multi oder Semantische Digitale Bibliotheken wie z. B. die Deutsche Digitale Bibliothek, sind ebenfalls eine Hybridstruktur zwischen Linear Multi und Social Web. Allerdings werden hier nur diejenigen Inhalte aus dem Social Web zugelassen, die durch ein Expertengremium auch für zulässig erklärt werden. Der Experte bestimmt, welche Inhalte in das geschlossene Expertensystem übernommen werden. Die Plattformen werden durch Experten gebaut. Eine Mitbestimmung der Benutzer ist indirekt möglich.

### **4 Diskussion / Interpretation**

Zentrale Mehrwertprozesse durch Bibliotheken können durch den Prozess Ordnung schaffen entstehen. Innerhalb dieser Prozesse werden Wissensarchitekturen erzeugt. Es wirken unterschiedliche Informationsarchitekturen bei der Gestaltung von tieferen Ordnungsstrukturen mit. Die tradierten Märkte der nichtkommerziellen Wissenschaftskommunikation werden durch die öffentlichen, sich selbst organisierenden, nichtkommerziellen Wissenschaftsmärkte erweitert.<sup>17</sup>

In Zukunft ist eine Verschmelzung der nicht-kommerziellen Wissenschaftsmärkte mit den kommerziellen Informationsmärkten wahrscheinlich.

---

<sup>17</sup> Vgl. Sorge, Stefanie (2012), Semantische Digitale Bibliotheken, 12ff.

Hierfür gilt es, die öffentlich erzeugten Informationen zu schützen und unter Lizenzen zu stellen, welche die Nutzung der Informationen nur unter gleichen Bedingungen erlauben. Die tradierten Gedächtnis- und Kultureinrichtungen befinden sich in einem Wettbewerbsumfeld. Bisher ist nicht hinreichend geklärt, wo die Qualitätsdiskussion in den Computernetzwerken stattfinden soll. Hier könnten die tradierten Experten der Bibliotheken, Dokumentation, den Archiven und Museen aktiver werden und Überlegungen zu neuen Dienstleistungen anregen.

Fachwissen und Expertentum werden für die Ordnung der Informationsflut benötigt. Die Qualität der Informationen kann durch bessere Ordnungstiefen durch Ordnen, Strukturieren, Selektieren, Aggregieren und Kontextualisieren erzeugt werden. Ob gänzlich auf intellektuelle Erschließung verzichtet werden kann, darf bezweifelt werden.

Ein Nachteil aller geschlossenen Systeme ist die Abschottung von der sozialen Realität. Die soziale Realität ist die Gesamtheit der geteilten Wirklichkeitskonstruktionen. Ein geschlossenes System nimmt die soziale Realität im Außen nicht wahr, da es ständig um sich selbst kreist. Auch wenn ein geschlossenes System sehr viele Informationen zu aggregieren vermag, ist es nicht davor geschützt, durch die Vernachlässigung der sozialen Realität, Fehleinschätzungen zu begehen. So kann es z. B. zur Vernachlässigung von bestimmten Suchtermini kommen oder Trends werden zwar erkannt, können jedoch nicht in eine Handlung übersetzt werden.

Daher ist es wichtig das Soziale in die zukünftigen semantischen Bibliotheken zu integrieren, damit Innovation und die Integration der sozialen Realität gelingt. Hierfür sind die Wahrnehmung der sozialen Felder<sup>18</sup> und die Fähigkeit wichtig, die Aufmerksamkeit auf diese sozialen Felder lenken zu können. Soziale Felder können durch lebende Systeme über Fühl-, Denk- und Verhaltensprogramme wahrgenommen werden. Daher ist der Mensch die strategisch wichtige Inputvariable in Computernetzwerken.

---

<sup>18</sup> Vgl. Scharmer, Otto C. (2011): Theorie U, 127.

## Literaturverzeichnis

- Baecker, Dirk*, Sternstunde Philosophie: Die Krisen der Computergesellschaft, in: <http://www.youtube.com/watch?feature=endscreen&NR=1&v=Wep6R-IOW-E>, 22.08.2012.
- Belliger, André/Krieger, David J.* (Hrsg.), *Antology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, Bielefeld (Transcript) 2006.
- Ciampi, Luc*, *Die emotionalen Grundlagen des Denkens. Entwurf einer fraktalen Affektlogik*, Göttingen (Vandenhoeck & Ruprecht) 3. Aufl. 2005.
- DGI e. V.*, Positionspapier. Information und Wissen in globalen Netzen nutzbar machen, (Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V.)  
in: <http://www.dgi-info.de/images/PDF/Positionspapier.pdf>, 08.08.2012.
- Hejl, Peter M.*, Konstruktion der sozialen Konstruktion. Grundlinien einer konstruktivistischen Sozialtheorie, in: *Schmidt, Siegfried J. (Hrsg.)*, *Der Diskurs des Radikalen Konstruktivismus*, Frankfurt am Main (Suhrkamp) 1987, 303-339.
- Hejl, Peter M.*, Die Entwicklung der Organisation von Sozialsystemen und ihr Beitrag zum Systemverhalten, in: *Rusch, Gebhard/Schmidt, Siegfried J. (Hrsg.)*, *Konstruktivismus und Sozialtheorie*, Frankfurt am Main (Suhrkamp) 1994, 109-132.
- Latour, Bruno*, *Die Hoffnung der Pandora*, Frankfurt am Main (Suhrkamp) 2002.
- Latour, Bruno*, Über technische Vermittlung. Philosophie, Soziologie und Genealogie, in: *Belliger, André/Krieger, David J.* (Hrsg.), *Antology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, Bielefeld (Transcript) 2006, 483-528.
- Luhmann, Niklas*, *Soziale Systeme. Grundriß einer allgemeinen Theorie*, Frankfurt am Main (Suhrkamp) 1984, 168.
- McLuhan, Marshall*, *Understanding media, the extensions of man*, London (Routledge & Paul) 1964.
- Maturana, Humberto R./Varela, Francisco J.*, *Autopoiesis and cognition. The realization of the living*, Dordrecht (Reidel) 1980.
- Rusch, Gebhard/Schmidt, Siegfried J. (Hrsg.)*, *Konstruktivismus und Sozialtheorie*, Frankfurt am Main (Suhrkamp) 1994.
- Scharmer, Otto C.*, *Theorie U – Von der Zukunft her führen. Presencing als soziale Technik*, Carl-Auer (Heidelberg) 2. erw. Aufl. 2011.
- Schmidt, Siegfried J. (Hrsg.)*, *Der Diskurs des Radikalen Konstruktivismus*, Frankfurt am Main (Suhrkamp) 1987.
- Sorge, Stefanie*, *Semantische Digitale Bibliotheken : Welchen Mehrwert produzieren Bibliotheken als Gestalter innovativer, semantisch basierter Wissensräume?* Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin) 2011.  
<http://edoc.hu-berlin.de/series/berliner-handreichungen/2011-312/PDF/312.pdf>, 08.08.2012.
- Sorge, Stefanie*, *Angst und Leistung in Teams der Medienwirtschaft, Systemtheorie. Marktanforderungen. Wirkfaktoren*, Wiesbaden (Springer Gabler) 2012.

---

# **Vernetztes Wissen – Systeme**



## **Metadaten aus der Cloud: Technologien und Anwendungsfälle der Medienerschließung mittels Software as a Service**

Michael Eble, Sebastian Kirch

Fraunhofer-Institut IAIS, Sankt Augustin

### **Zusammenfassung**

Metadaten sind der Schlüssel, um große Medienbestände besser zu nutzen. Ihre Erzeugung ist dabei meist mit großem finanziellem und personellem Aufwand verbunden. Doch mit dem Wachsen digitaler Medienarchive und dem Entstehen neuer IT-Paradigmen wie dem Cloud Computing entstehen auch neue Lösungen zur Medienerschließung: Durch Software as a Service werden technische, organisatorische und finanzielle Zugangshürden verringert, da für Archivbetreiber das Anschaffen und Betreiben einer eigenen Infrastruktur zur Medienerschließung entfällt. Vor allem in den USA gibt es bereits mehrere Anbieter, die sich auf das Erzeugen von Metadaten als Service spezialisiert haben. Auch in Deutschland entstehen solche Angebote. Ein Beispiel dafür ist die CONTENTUS-Diensteplattform aus dem THESEUS-Forschungsprogramm. Ein Nutzen solcher Angebote ist es, dass automatische und manuelle Erschließungsverfahren verschränkt werden können, um ein qualitativ hochwertiges Ergebnis zu erzielen.

Dieser Beitrag stellt relevante Konzepte, Technologien sowie Anwendungsfälle vor und gibt einen Ausblick auf zukünftige Entwicklungen.

### **Abstract**

Metadata is the key to better exploit large media collections. However the generation of metadata is usually associated with enormous financial investments and human time and effort. But due to the growth of digital media archives and the formation of new paradigms in the IT industry like cloud computing, new solutions for content enrichment: The main advantage of these new using software as a service business models is the reduction of technical, organizational and financial barriers since archive owners are released from procurement and operation of expensive media enrichment infrastructures. Primarily in the USA several metadata service providers already exist. In Germany equivalent offers begin to evolve, e. g. the CONTENTUS Service Platform for Content Enrichment. One benefit of such offers is the combination of automatic and manual enrichment techniques to generate premium-quality results.

This article presents relevant concepts, technologies as well as use cases and gives a perspective on future trends.



## **1. Einleitung und Problemaufriss**

Ziel des vorliegenden Beitrags ist es, aktuelle Möglichkeiten und Grenzen sowie Angebote und Entwicklungsperspektiven der Medienerschließung mittels Software as a Service vorzustellen. Dabei geht es im Wesentlichen darum, dass Hard- und Software-Ressourcen zur Umsetzung von Erschließungsprojekten nicht lokal vorgehalten werden, sondern über Netzwerke genutzt werden.

Der Überblicksbeitrag stellt dazu zunächst das Konzept von Metadaten aus der Cloud sowie eine Auswahl relevanter Technologien vor. Anschließend bespricht er verschiedene Anwendungsfälle und gibt einen Ausblick über weitere Entwicklungen. Er basiert auf Marktstudien sowie Projektarbeiten zur Medienerschließung via Cloud Computing und aktualisiert Inhalte einer früheren Veröffentlichung (Eble/Kirch 2012).

### *Herausforderungen in Projekten zur Medienerschließung*

Das Erschließen von Medienbeständen ist eine zentrale Grundlage, um ihre Inhalte kontextualisieren und nutzen zu können. Das Erzeugen von reichhaltigen Metadaten gehört daher zu einer wesentlichen Tätigkeit von Bibliotheken und Archiven in öffentlichen Einrichtungen und Unternehmen. Dazu werden manuelle und automatische Verfahren kombiniert und deren Ergebnisse miteinander verbunden. Insbesondere zur effizienten automatischen Analyse von großen, digitalen Audio-/Video- sowie Bild-/Dokumentenbeständen sind leistungsfähige Hard- und Softwarearchitekturen notwendig. Diese sind meist mit hohen fixen Kosten verbunden. Weiterhin stehen Organisationen vor der Herausforderung, Erschließungsprojekte in engen Kosten- und Zeitrahmen zu realisieren bzw. die Erschließung von großen Beständen überhaupt erst realisieren zu können.

### *Chancen und Risiken von Medienerschließung via Cloud Computing*

Ansätze des Cloud Computings propagieren demgegenüber nun, Dienste zur Medienerschließung so bereitzustellen, dass eine Verwendung bei weitgehend variablen Kosten möglich wird: Der Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (BITKOM) definiert als Cloud Computing „eine Form der Bereitstellung von gemeinsam nutzbaren und flexibel skalierbaren IT-Leistungen durch nicht fest zugeordnete IT-Ressourcen über Netze“, die eine Umverteilung von Investitions- zu Betriebsaufwänden ermögliche (BITKOM 2010,

15). Generell werden mehrere aufeinander aufsetzende Ebenen voneinander abgegrenzt: Von Infrastrukturen (IaaS) über Plattformen (PaaS) und Software (SaaS) bis zu Mitarbeitern (HuaaS) und Daten (DaaS) sind grundsätzlich eine Reihe von IT-basierten Leistungen via Cloud Computing nutzbar (vgl. BITKOM 2010; Lenk et al. 2009; Pellegrini 2012; siehe Abbildung 1). Medienerschließung via Cloud Computing meint dementsprechend die Aufbereitung, strukturelle und inhaltliche Erschließung sowie semantische Anreicherung von digitalen Medienbeständen mittels z. B. Software as a Service.

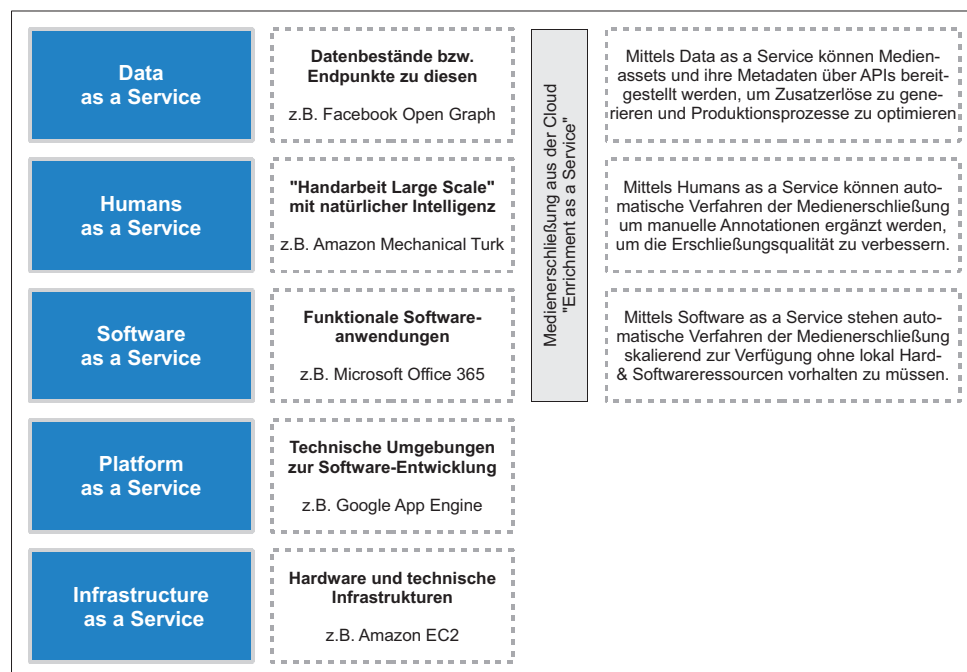


Abbildung 1: Ebenen des Cloud Computings und ihr Bezug zur Medienerschließung.  
Quelle: Eble/Kirch 2012

Wertet man existierende Studien aus, dann lassen sich unterschiedliche Chancen und Risiken von Cloud Computing im Allgemeinen sowie der darüber durchgeführten Medienserschließung im Speziellen identifizieren (s. Holtkamp 2010, PwC 2011 und Weiner 2011; vgl. Eble/Kirch 2012): Für Bibliotheken und Archive sind damit zunächst eine Reihe von potentiellen Chancen offensichtlich: Sie erhalten die Möglichkeit, Software as a Service über Dienstplattformen bedarfsgerecht zu nutzen – ohne zur Projektdurchführung eigene Hardware- und Software-Infrastrukturen lokal vorhalten zu müssen. Über Schnittstellen sollen Verfahren und

Ergebnisdaten in eigene Anwendungen und Archiv-Systeme integriert werden (können). Dadurch können die finanziellen, organisatorischen und technischen Hürden sinken, um automatische Verfahren der Medienschließung zu nutzen. In zeitlicher Hinsicht können einerseits verkürzte Zeiten bis zum Projekt-Setup erreicht werden und andererseits können verkürzte Vertragslaufzeiten realisiert werden. Daneben kann prinzipiell die Vergleichbarkeit von Anbietern bzw. ihren Leistungen steigen sowie die Interoperabilität zwischen Lösungen verschiedener Anbieter verbessert werden.

Diesen Chancen steht jedoch eine ganze Reihe von nicht zu unterschätzenden Risiken gegenüber, die den Chancen teilweise diametral entgegenstehen: Erstens befindet sich der Markt für Medienschließung via Cloud Computing noch in einer Frühphase und ist entsprechend unübersichtlich. Gleichzeitig besteht Unsicherheit im Hinblick auf die tatsächliche Leistungsfähigkeit der Angebote sowie hinsichtlich ihrer Sicherheit bei der Datenspeicherung und -verarbeitung. Hinzu kommen Risiken hinsichtlich der tatsächlichen Interoperabilität und der damit verbundenen Bindung an bestimmte Anbieter oder Lösungen (Vendor-Lock-In). Schließlich – und das ist mit der wesentlichste Aspekt – sind nur sehr wenige Anwenderberichte und dokumentierte Best Practices verfügbar.

Um das Potential von Medienschließung via Cloud Computing besser einschätzen zu können werden ihre wesentlichen Aspekte nachfolgend systematisiert, um dann Technologien und Anbieter sowie Anwendungsfälle zu besprechen.

## **2. Metadaten aus der Cloud**

Ausgangspunkt der Digitalisierung und Erschließung von Medienbeständen können unterschiedliche Ziele sein wie z. B. das Sichern und Bewahren von Kultur- und Mediengütern, das Finden und Vernetzen von digitalen Inhalten oder das Bündeln und Verwerten von digitalen Produkten (siehe Abbildung 2). In jedem Fall sind dann u. a. identifizierende, beschreibende, struktur- und relationsbezogene sowie administrative und technische Metadaten erforderlich, so dass das Erzeugen von reichhaltigen Metadaten zu einer wesentlichen und sich wiederholenden Maßnahme wird. Als Metadaten aus der Cloud stehen die Ergebnisse der Medienschließung mittels Software as a Service entsprechend zur Verwendung zur Verfügung.

Das lässt sich am Beispiel der Dokumentenerschließung konkretisieren: Mittels automatischer Verfahren findet im ersten Schritt ein *Aufbereiten der Mediendaten* statt, wodurch die Scans von Zeitungs- oder Buchseiten geschärft und entzerrt werden, Alterungseffekte entwerft und durch eine Rotationskorrektur angemessen ausgerichtet werden. Im zweiten Schritt wird ein *Erkennen von Strukturen* durchgeführt, wobei Buch- oder Zeitungsseiten in ihre logischen Einheiten unterteilt und zusammengehörende Zeitungsartikel identifiziert werden, so dass XML-Beschreibungen der logischen und physischen Struktur von Dokumenten vorliegen. Zudem wird z. B. die jeweilige Titelseite einer Zeitung oder eines Buches automatisch erkannt und extrahiert, so dass sie als Vorschau o. ä. genutzt werden kann. Im dritten Schritt wird ein *Erschließen und Anreichern von Inhalten* umgesetzt, so dass mittels Texterkennung (OCR) Bild- in maschinenlesbare Textdaten umgewandelt werden und die pixelgenaue Position von Texten, Grafiken, Bildern, Tabellen etc. sowie logische Bezeichner für jedes Element eines Digitalisats vorliegen. Aus den maschinenlesbaren Texten können dann z. B. benannte Entitäten wie Personen, Organisationen und Orte extrahiert, identifiziert und mit Datensätze wie Biografien oder Dossiers verlinkt werden.

In der Verschränkung mit den genannten automatischen Verfahren können manuelle Schritte umgesetzt werden. Das kann erstens ein *Trainieren von Algorithmen* sein – z. B. um für einen Trainingskorpus Dokumente, Entitäten oder andere Daten zu klassifizieren, der als Referenzpunkt für maschinelle Lernverfahren dient. Zweitens handelt es sich um ein *Optimieren von Metadaten* wie etwa das Korrigieren von Fehlern, die im Zusammenhang mit automatischen Verfahren zur strukturellen und inhaltlichen Erschließung entstanden sind. Drittens kann dies auch ein *Ergänzen von Metadaten* sein, wenn manuell Informationen zu Personen oder Links zu biografischen Informationen und Standorten hinzugeführt werden.

Die skizzierten Schritte kommen selbstverständlich nicht nur für die Dokumentenerschließung zum Einsatz, sondern in Varianten auch für die Erschließung von Audio- und Videobeständen. Dabei handelt es sich hinsichtlich des Erkennens von Strukturen z. B. um Audio- bzw. Videosegmentierung oder um Sprechergruppierung. Hinsichtlich des Erschließens von Inhalten zählt dazu die Sprach- und Sprechererkennung sowie die Identifikation von Gesichtern. Auch können sich automatische und manuelle Verfahren ergänzen.

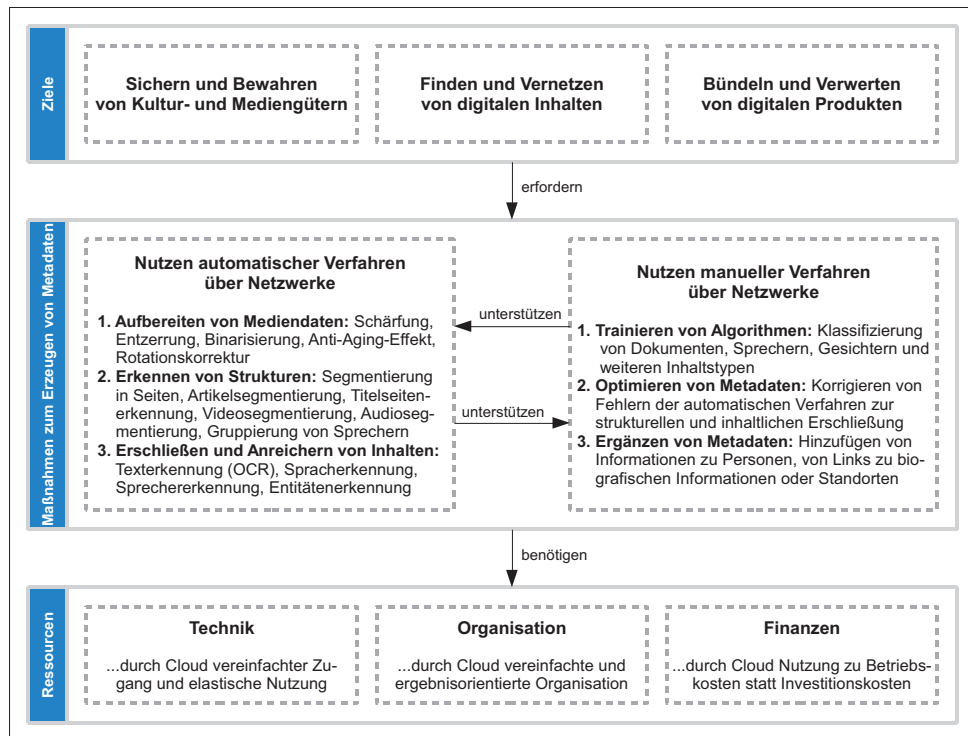


Abbildung 2: Metadaten aus der Cloud: Ziele, Maßnahmen und Ressourcen aus der Sicht von Nutzern

Die Notwendigkeit bzw. der Nutzen der Verschränkung von manuellen und automatischen Verfahren hängt damit zusammen, dass einerseits Projekte mit großen Medienbeständen mit rein manueller Arbeit nicht auf wirtschaftliche Weise erschlossen werden können und andererseits automatische Verfahren teilweise trainiert werden müssen und nicht immer vollständig zufriedenstellende Ergebnisse liefern können. Wesentlich ist, dass bei der Medienererschließung via Cloud Computing verschiedene Formen genutzt und ergänzt werden können, indem die jeweiligen Leistungen z. B. als Web-Services bereitstehen und entsprechend Internet-basiert genutzt werden können.

Ausgehend vom skizzierten Konzept wird nachfolgend ein Überblick über Technologien und Anbieter für die Cloud-basierte Erschließung von digitalen Bild-, Audio-, Video- und Textdaten gegeben.

### 3. Überblick über Anbieter und Technologien

Die tabellarische Aufstellung enthält mehrere kommerzielle Anbieter und ihre Technologien, die hinsichtlich ihrer Leistungsversprechen, der verarbeitbaren

Datentypen und der unterstützten Sprachen gegenübergestellt sind. Die Angaben entstammen den Webseiten der jeweiligen Anbieter.

**Tabelle 1: Anbieter und Technologien zur Erschließung von Digitalisaten/Bilddaten**

<i>Bilddaten von Digitalisaten</i>				
	<b>Anbieter Produkt</b>	<b>Leistungen der Technologien</b>	<b>Verarbeitbare Datentypen</b>	<b>Unterstützte Sprachen</b>
1	Abbyy  Cloud OCR SDK	Automatische Texterkennung (OCR), um Digitalisate strukturell und inhaltlich zu erschließen und letztlich durchsuchbar zu machen; Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell	Digitalisate von Dokumenten als Bilddaten (Scans von Zeitungen, Büchern etc.; Fotos von einzelnen Zeitschriften- oder Zeitungs-seiten, von Büchern) als TIFF, PDF oder in anderen Formaten	198 Sprachen (Deutsch, Englisch, Französisch, Spanisch u. a.)
2	Kooaba  Technology Plattform	Automatische Mustererkennung, um Links zwischen Bildinhalten von Dokumenten (z. B. Werbeanzeige) und ergänzenden Inhalten aus dem Internet (z. B. Shop-Angebote) herzustellen; Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell		Sprachen-unabhängig
3	Nuance  OmniPage Cloud Service	Automatische Texterkennung (OCR), um Digitalisate strukturell und inhaltlich zu erschließen und letztlich durchsuchbar zu machen; Bereitstellung über Web-Schnittstellen (APIs); derzeit keine Preis- und Abrechnungsangaben		127 Sprachen (Deutsch, Englisch, Französisch, Spanisch u. a.)
4	Savetz Publishing  BetterOCR	Automatische Texterkennung (OCR) und manuelle Nach-bearbeitung durch Mitarbeiter des Anbieters, um Digitalisate durchsuchbar zu machen; Bereitstellung über eine Web-Oberfläche		Englisch
5	Smart Mobile Software  Online OCR API	Automatische Texterkennung (OCR), um Digitalisate durchsuchbar zu machen; zielt auf Entwickler von mobilen Anwendungen (Apps); Bereitstellung; Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell		38 Sprachen (Arabisch, Bulgarisch, Chinesisch, Deutsch, Englisch u. a.)
6	WiseTrend  OCR Cloud 2.0	Automatische Texterkennung (OCR), um Digitalisate durchsuchbar zu machen; zielt auf Entwickler von mobilen Anwendungen (Apps); Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell		Über 200 Sprachen (Lateinische Schriften, Kyrillische Schriften, Chinesisch etc.)

Tabelle 2: Anbieter und Technologien zur Erschließung von Audio- und Videodaten

<i>Audio- und Videodaten (gesprochene Sprache)</i>				
	<b>Anbieter Produkt</b>	<b>Leistungen der Technologien</b>	<b>Verarbeitbare Datentypen</b>	<b>Unterstützte Sprachen</b>
1	3Play Media	Kombination von automatischen und manuellen Verfahren, um gesprochene Videoinhalte in zeitsynchrone Textinhalte umzuwandeln; u. a. zur Suchmaschinenoptimierung (Video-SEO) und für Untertitel; Textinhalte auch in anderen Sprachen durch Übersetzung der gesprochenen Inhalte; Bereitstellung über Web-Schnittstellen (APIs)	Digitale Audio-Video-Daten mit gesprochener Sprache (z. B. Nachrichtensendungen, Dokumentationen, Reportagen oder Podcasts)	Englisch, Französisch, Spanisch
2	Greenbutton InCus	Bereitstellen von automatischen Verfahren über Web-Schnittstellen (API), um gesprochene Videoinhalte in zeitsynchrone Textinhalte umzuwandeln; u. a. zur Suchmaschinenoptimierung (Video-SEO) und für Untertitel		Englisch
3	SpeakerText	Kombination von automatischen und manuellem Verfahren, um gesprochene Videoinhalte in zeitsynchrone Textinhalte umzuwandeln		Englisch
4	VeeSeo	Automatische Verfahren, um gesprochene Videoinhalte in zeitsynchrone Textinhalte umzuwandeln; sprecherunabhängige Spracherkennung und semantische Anreicherung; Bereitstellung über Web-Schnittstellen (APIs)		Deutsch

Tabelle 3: Anbieter und Technologien zur semantischen Anreicherung von Textdaten

<i>Maschinenlesbare Textdaten</i>				
	Anbieter Produkt	Leistungen der Technologien	Verarbeitbare Datentypen	Unterstützte Sprachen
1	Evri Developer API	Automatische Verfahren u. a. für Named Entity Extraction, Concept Tagging und Textklassifikation, um Textdaten inhaltlich zu erschließen und semantisch anzureichern; Bereitstellung über Web-Schnittstellen (APIs)	Maschinenlesbare Textdaten als HTML, TXT, XML oder in anderen Formaten	Englisch
2	Open Amplify Insights API	Automatische Verfahren u. a. für Named Entity Extraction, Concept Tagging, Textklassifikation und Sprachdetektion, um Textdaten inhaltlich zu erschließen und semantisch anzureichern; Bereitstellung über Web-Schnittstellen (APIs)		Englisch
3	Orchestr8 AlcheMyAPI	Automatische Verfahren u. a. für Named Entity Extraction, Concept Tagging, Textklassifikation und Sprachdetektion, um Textdaten inhaltlich zu erschließen und semantisch anzureichern; Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell		Englisch, Französisch, Deutsch, Italienisch, Spanisch u. a.
4	Textwise API	Automatische Verfahren u. a. für Named Entity Extraction und Textklassifikation, um Textdaten inhaltlich zu erschließen und semantisch anzureichern; Bereitstellung über Web-Schnittstellen (APIs)		Englisch
5	Thomson Calais	Automatische Verfahren u. a. für Named Entity Extraction und Textklassifikation, um Textdaten inhaltlich zu erschließen und semantisch anzureichern; Bereitstellung über Web-Schnittstellen (APIs) und in einem mengenbasierten Abrechnungsmodell		Englisch, Französisch, Spanisch

Insgesamt steht eine Reihe von Werkzeugen als Cloud-basierte Modelle zur Verfügung, die eine Erschließung und Anreicherung von Medienbeständen in Form von Bild-, Audio-, Video- und Textdaten ermöglichen. Die Angebote sind dabei noch in frühen Stadien und teilweise noch in Beta- oder Evaluierungsphasen – so sind z. B. der OmniPage Cloud Service von Nuance und das Abbyy Cloud OCR SDK im ersten Halbjahr 2012 gestartet. Dabei unterstützen die meisten Angebote englischsprachige Datenbestände, während deutschsprachige Inhalte weniger



umfassend unterstützt werden. Hinzu kommt, dass eine systematische Verschränkung von automatischen und manuellen Verfahren hier nur wenig beobachtet wurde.

Darüber hinaus ist zur Kombination von z. B. der Verarbeitung von Digitalisaten und der anschließenden semantischen Anreicherung der Textdaten eine Verschaltung verschiedener Anbieter erforderlich. Ein dazu bereits existierendes Angebot ist die MediaCloud des US-amerikanischen Anbieters Ramp. Nach Angaben des Betreibers können damit einerseits mehrere Datentypen verarbeitet werden und andererseits werden Daten sowohl strukturell und inhaltlich erschlossen als auch semantisch angereichert.

#### 4. Die CONTENTUS-Diensteplattform und ihre Anwendungen

Die Anwendungsforschung im Bereich von Bibliotheken und Archiven beschäftigt sich unter verschiedenen Fragestellungen ebenfalls mit der Digitalisierung, Erschließung und Verwertung von Medienbeständen. In dem vom Bundesministerium für Wirtschaft und Technologie (BMWi) geförderten Forschungsprogramm THESEUS adressiert das Teilprojekt CONTENTUS entsprechende Fragestellungen. Der Ausgangspunkt dabei ist, dass Kultur- und Medienorganisationen vor mehreren typischen Problemen stehen, die sich vom Zerfall von Archivgut über unzureichende Erschließung und fehlende Sinnzusammenhänge bis zu unklaren Verwertungsperspektiven erstrecken. Ziel des Projekts ist es daher, für alle o. g. Prozessschritte und Medientypen Lösungen mit einem möglichst hohen Automatisierungsgrad zu entwickeln (siehe Abbildung 3).

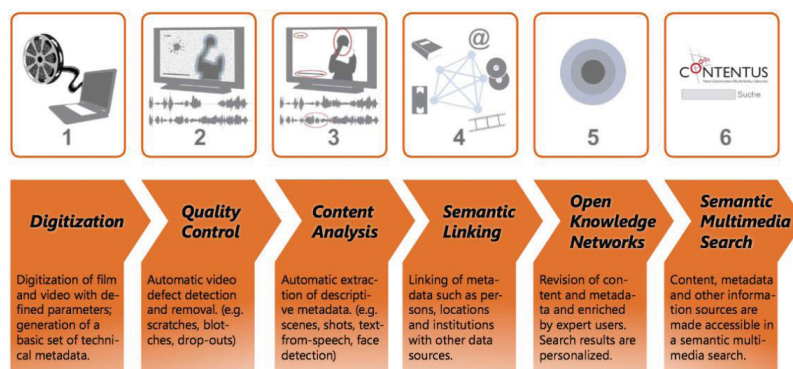


Abbildung 3: Die CONTENTUS-Prozesskette der Medienererschließung

Wesentlich ist dabei, dass insbesondere solche Lösungen bereitstehen, die auf die Bedarfe von Kultureinrichtungen wie Bibliotheken und Archive zugeschnitten sind, indem sie den Umgang mit schwierigen Digitalisaten von Zeitungen und Büchern adressieren. Ein Beispiel dafür ist das Aufbereiten von Digitalisaten mit schlechtem Kontrastverhältnis zwischen Hintergrund und Text sowie von Ormig-Kopien.

Ein zentrales Ergebnis des Projekts ist die CONTENTUS-Diensteplattform. Die Plattform bündelt eine Reihe von Verfahren zur automatischen Medienererschließung und stellt sie zur Nutzung über Web-Schnittstellen und -Oberflächen bereit. Zu den Verfahren zählt erstens das Aufbereiten von digitalisierten Print-Dokumenten (z. B. Schärfung, Entzerrung, Binarisierung, Entfernen von Alterungseffekten). Zweitens handelt es sich um die strukturelle Erschließung (z. B. Seiten- und Artikelsegmentierung) und drittens um die inhaltliche Erschließung (z. B. Texterkennung, OCR) und semantische Anreicherung (z. B. Erkennung von Entitäten). Schließlich ist ein Export der Digitalisate und Metadaten in Formaten wie PDF/A, METS, MODS, ALTO u. a. möglich.

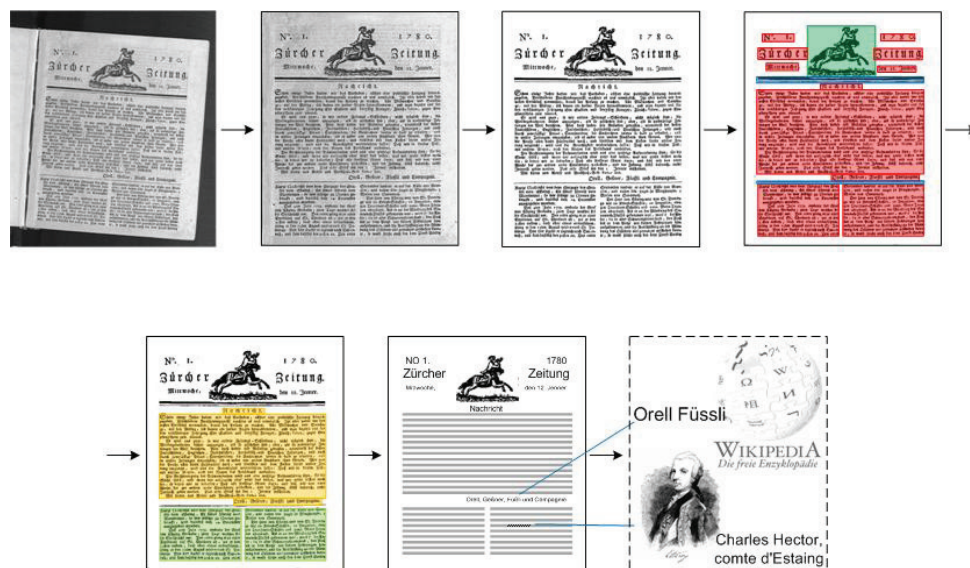


Abbildung 4: Prozess zur Aufbereitung und Erschließung von Zeitungsdigitalisaten in CONTENTUS

Derzeit befindet sich die CONTENTUS-Diensteplattform in einer Invited-Beta-Phase und wird gemeinsam mit Testern aus verschiedenen Organisationen weiterentwickelt und verbessert. Ein erster Anwendungsfall ist dabei die Verwendung der Plattform durch die Deutsche Digitale Bibliothek und ihre beteiligten Einrichtungen:

Kultureinrichtungen und Digitalisierungsdienstleister nutzen die im Rahmen von CONTENTUS entwickelte Plattform, um Digitalisate aufzubereiten sowie strukturell und inhaltlich zu erschließen und tragen mit ihrem Feedback gleichzeitig zur Verbesserung bei. Die Evaluierungsphase läuft derzeit noch und die Ergebnisse fließen anschließend in die weitere Produktentwicklung ein. Ein weiterer Anwendungsfall ist die Erschließung der Zeitschrift EMMA: Der Bonner Digitalisierungsdienstleister ImageWare hat zusammen mit Fraunhofer IAIS und dem Hochschulbibliothekszenrum NRW den gesamten Bestand der Zeitschrift EMMA online zugänglich gemacht. Mit dem neuen Produkt EMMAdigital stehen alle bisherigen Ausgaben seit 1977 digitalisiert und kostenlos über das Internet zur Verfügung. Ein dritter Anwendungsfall ist der Aufbau eines Artikelarchivs für die historischen Ausgaben von Neues Deutschland. Hierzu wurden zunächst 15.000 Ausgaben der Zeitung über die Dienstplattform strukturell und inhaltlich erschlossen. Da das geplante Artikelarchiv eine möglichst hohe Qualität bieten soll, hat der Berliner Archivdienstleister ArchivInForm die Ergebnisse der automatischen Erschließungsverfahren als Ausgangsbasis genutzt und sie mit einer Korrektur-Anwendung der Plattform manuell aufbereitet.

## **5. Ausblick und weitere Entwicklung**

Der Überblick über bereits am Markt verfügbare Produkte hat gezeigt, dass zur Umsetzung von Projekten eine Medienerschließung via Cloud Computing auf der technischen Ebene grundsätzlich realisierbar ist. Es ist davon auszugehen, dass in der weiteren Marktentwicklung zunehmend mehr Anbieter ihre Verfahren zum Aufbereiten, Erschließen und Anreichern von Medienbeständen als Cloud-Lösungen bereitstellen werden. Offen ist derzeit jedoch, welche Best Practices dabei entstehen werden, was den Umgang mit „schwierigen“ Digitalisaten und die Verbindung von Produkten verschiedener Anbieter zu einer integrierten Verarbeitungskette betrifft. Hierzu soll eine Evaluierungsstudie umgesetzt werden, die erstens Anbieter mit einem definierten Set von Testdaten (z. B. Digitalisate von Zeitungen und Büchern) vergleicht und zweitens die Orchestrierbarkeit der APIs verschiedener Anbietern adressiert. Auf diese Weise würde für Digitalisierungsprojekte eine belastbare Einschätzung zum Status quo von Metadaten aus der Cloud geschaffen.

### Literaturverzeichnis

BITKOM (2010): Cloud Computing – Was Entscheider wissen müssen. Ein ganzheitlicher Blick über die Technik hinaus. URL:  
[http://www.bitkom.org/60376.aspx?url=BITKOM\\_Leitfaden\\_Cloud\\_Computing-Was\\_Entscheider\\_wissen\\_muessen.pdf&mode=0&b=Publikationen](http://www.bitkom.org/60376.aspx?url=BITKOM_Leitfaden_Cloud_Computing-Was_Entscheider_wissen_muessen.pdf&mode=0&b=Publikationen)

Eble, Michael; Kirch, Sebastian (2012): Metadaten aus der Cloud: Erschließung von Medienarchiven über Dienstplattformen und Software as a Service. In: Info 7 - Medien | Archive | Information (2012).

Holtkamp, Bernd (2010): Cloud Computing für den Mittelstand am Beispiel der Logistikbranche. Dortmund: Fraunhofer ISST.

Lenk, Alexander; Klems, Markus; Nimis, Jens; Tai, Stefan; Sandholm, Thomas (2009): What's inside the Cloud? An architectural map of the Cloud landscape. In: Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing (CLOUD '09). IEEE Computer Society, Washington, DC, USA, S. 23-31.

Pellegrini, Tassilo (2012): Quality-of-Service in der Netzbewirtschaftung – Effekte einer eingeschränkten Netzneutralität auf Cloud-basierte Geschäftsmodelle. In: Krone, Jan; Pellegrini, Tassilo (Hg.). Netzneutralität und Netzbewirtschaftung. Baden-Baden: Nomos Verlag, S. 117-137.

PricewaterhouseCoopers (2011): Cloud Computing im Mittelstand. Erfahrungen, Nutzen und Herausforderungen.

Weiner, Nico; Renner, Thomas; Kett, Holger (2010): Geschäftsmodelle im „Internet der Dienste“. Aktueller Stand in Forschung und Praxis. Stuttgart: Fraunhofer IAO.

---

## **Open-Access-Netzwerk: Semantische Anreicherung und Vernetzung von wissenschaftlichen OA-Inhalten**

Sammy David<sup>1</sup>, Julia Iwanowa<sup>2</sup>, Maxi Kindling<sup>2</sup>, Michaela Voigt<sup>2</sup>

<sup>1</sup> Computer- und Medienservice der Humboldt-Universität zu Berlin

<sup>2</sup> Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin

### **Zusammenfassung**

Der Beitrag bietet einen Überblick über das Projekt OA-Netzwerk, den gleichnamigen Dienst und Datenraum deutscher Repositorien sowie die Vorgehensweise bei der Publikation des OA-Netzwerk-Datenraums als Linked Open Data (LOD). Die Projektergebnisse werden dabei in den Kontext der aktuellen Entwicklungen im Bereich des Semantic Web unter besonderer Berücksichtigung der LOD-Veröffentlichung bibliografischer Daten in Deutschland eingeordnet. Die Herausforderungen, die die technische Vernetzung und der (nationale) Nachweis von wissenschaftlichen Inhalten (auch Enhanced Publications, Forschungsdaten) als Linked Open Data mit sich bringen, stehen dabei im Vordergrund.

### **Abstract**

This paper gives an overview on the project Open Access Network as well as the same-titled services and data aggregation of German repositories. Further it describes the experimental workflow for contextualising and interlinking strategies with linked open data technologies and standards for the metadata of open access scholarly publications in the infrastructure of Open Access Network. Therefore the project and its results are outlined within in the context of Semantic Web developments and technologies.

OA-Network is a project initiated by the German Initiative for Network Information (DINI) and was founded by the German Research Foundation conducted jointly by the Humboldt University of Berlin (Computer and Media Services), the Lower Saxony State and University Library and the University of Osnabrueck (Department of Mathematics and computer science). The goal of the project was to provide a network of certified open access repositories in Germany to build up and thereby strengthen the open-access infrastructure for German research.

### **Forschungsstand und Einordnung von OA-Netzwerk**

Weltweit gibt es derzeit mehr als 2000 digitale Repositorien für Textpublikationen (vgl. u. a. OpenDOAR), die Zahl der Repositorien für digitale Forschungsdaten ist nicht erfasst.<sup>1</sup> Auch in Deutschland haben Textrepositorien eine starke Verbreitung erlangt, wie die Anzahl der Repositorien verzeichnet, die bei der Deutschen Initiative für Netzwerkinformation e.V. aufgeführt werden (vgl. DINI-Liste der Repositorien, derzeit noch mit einem Zuschnitt auf Textrepositorien). Die in den Repositorien bereitgestellten Dokumente werden auf diese Weise für Fachcommunities und die globale Öffentlichkeit aufbereitet und sichtbar gemacht. Damit wird ein enormer Wissenspool zugänglich. In der Praxis bleiben die Dokumente in Textrepositorien dennoch häufig verborgen oder sind schwer auffindbar. Zwar ermöglichen Volltextsuchmaschinen wie Google oder Bing alternative Zugänge, mit denen sich eine Vielzahl an Informationen finden lässt. Jedoch gehen wissenschaftliche Dokumente ohne präzise gezielte Suchanfrage oft im Rauschen des World Wide Web unter. In der Regel fehlt bei allgemeinen Suchmaschinen die Möglichkeit eines konkreten Kontextfilters für die Suchergebnisse, wie er für die Wissenschaft sinnvoll wäre. Zudem erweist sich das getrennte Vorliegen von Daten in den so genannten "Datensilos" als nachteilig. Erst durch einen Zusammenschluss zu einer vernetzten Gesamtstruktur lässt sich das Potential umfassend ausschöpfen. Die physische Bereitstellung von Dokumenten und deren Erschließung durch Metadaten in Repositorien ist nur ein erster Schritt. Durch die Zusammenführung dieser Daten und ihre semantische Kontextualisierung, also das Inbeziehungsetzen dieser Daten aus den verschiedenen Silos, lässt sich die Sichtbarkeit deutlich steigern. Eine Vernetzung im Sinne des Semantic Web ermöglichte sowohl im Allgemeinen als auch mit Blick auf Open-Access-Netzwerk (kurz: OA-Netzwerk oder OAN) im Speziellen entsprechende positive Effekte.

Zu den Vorteilen von sog. Repositorien-Förderationen über offene Protokolle wie OAI-PMH gibt es bereits einige Arbeiten (vgl. bspw. Aschenbrenner et al., 2010 oder Müller et al., 2009). Der Ansatz des Projekts OA-Netzwerk geht in seiner zweiten Projektphase über die Abfrage und die Aggregation von Informationsressourcen (Metadaten) heterogener Datenquellen über OAI-PMH hinaus.

---

<sup>1</sup> Es gibt jedoch verschiedene Ansätze des Nachweises digitaler Repositorien für Forschungsdaten wie bspw. Databib ([www.databib.org](http://www.databib.org)) oder re3data.org ([www.re3data.org](http://www.re3data.org)).

Vielmehr sollen möglichst sämtliche Informationsressourcen eines solchen Aggregationsknotens zueinander in Beziehung gesetzt und damit kontextualisiert werden. Die aggregierten Daten sollen zudem mit externen Datenquellen verknüpft werden, um die Silobildung zu vermeiden. Als Informationsressourcen gelten im Folgenden bibliografische Daten und weitere Metadaten in Repositorien und Bibliothekskatalogen, die Publikationen, Forschungsdaten und weitere für Wissenschaft und Forschung relevante digitale und analoge Objekte beschreiben und identifizieren.

Das von der Deutschen Initiative für Netzwerkinformation e.V. (DINI) initiierte und durch die Deutsche Forschungsgemeinschaft (DFG) geförderte Projekt Open-Access-Netzwerk wurde in zwei Projektphasen (2007-2009 & 2009-2012) durch die Projektpartner der Humboldt-Universität zu Berlin, der Niedersächsischen Staats- und Universitätsbibliothek Göttingen sowie der Universität Osnabrück durchgeführt. Der Aufbau einer übergreifenden Infrastruktur für ein Netzwerk von Repositorien in Deutschland sollte die hiesige Open-Access-Infrastruktur an sich und damit die Sichtbarkeit der Publikationen der deutschen Forschungslandschaft stärken.

Die Grundlage der OA-Netzwerk-Infrastruktur und der mit OA-Netzwerk verbundenen oder kooperierenden Projekte sind die Standards des DINI-Zertifikats für Dokumenten- und Publikationsservices. Alle DINI-zertifizierten Repositorien sind derzeit in den Datenraum von OA-Netzwerk eingebunden, der die Metadaten der in den Repositorien vorgehaltenen Dokumente beinhaltet. Die mittels des OAI-PMH geharvesteten Metadaten werden in einer Kette von aufeinanderfolgenden Diensten aggregiert, aufbereitet und über eine Websuche als Recherchewerkzeug für EndnutzerInnen angebunden (vgl. OAN-Suche). Derzeit wird der Datenraum von OA-Netzwerk um die Repositorien erweitert, die noch kein DINI-Zertifikat besitzen.<sup>2</sup>

Zugleich werden die aggregierten Metadaten auch über standardisierte Schnittstellen wie einen SPARQL-Endpoint zur weiteren automatischen Verarbeitung nutzbar gemacht und als Open Data über die Anmeldung bei „The Data Hub“ (CKAN) angeboten (siehe auch: Linked Data Cloud). Die über OA-Netzwerk verfügbaren Informationsressourcen sind somit im Standardformat Resource Description

---

<sup>2</sup> Zum Zeitpunkt des Vortrags wird die Einbindung fortgeschritten sein.



Framework (RDF) für Dritte verfügbar und für die Anreicherung und Verlinkung externer Datenquellen bzw. eigener Bestände nutzbar. Die Verknüpfung der Metadaten mit anderen Datenquellen (z. B. die Autoreninformationen im OA-Netzwerk mit der Gemeinsamen Normdatei der Deutschen Nationalbibliothek) ist derzeit in Bearbeitung.

### OA-Netzwerk-Dienste und die technische Vernetzung von Repositorien

Im Rahmen von OA-Netzwerk wurde zunächst eine Basisinfrastruktur zum Harvesting der Open-Access-Dokumente und der dazugehörigen Metadaten aus DINI-zertifizierten Repositorien aufgebaut. Der so entstandene Datenraum umfasst 35 Repositorien mit knapp 150.000 nachgewiesenen Dokumenten.<sup>3</sup> Die aggregierten Metadaten werden zunächst harmonisiert. Neben einer Volltextsuche über die eigentlichen Textressourcen (Dokumente) sollte auch die Suche basierend auf den ausgewiesenen Metadatenfeldern beispielsweise für ein Browsing nach Dokumenttyp bzw. Fachklassifikation ermöglicht werden. Aufgrund der Heterogenität der in den angeschlossenen Repositorien verfügbaren Metadaten mussten jedoch zunächst Teile der Metadaten vereinheitlicht werden. So galt es beispielsweise, Feldinhalte einem einheitlichen Vokabular (z. B. DINI-Dokumenttypen) anzugleichen (siehe auch DINI-Zertifikat 2010).

Abbildung 1 illustriert die Auflösung dieser Heterogenität bzw. Mehrdeutigkeiten durch die Abbildung auf diskrete Werte.<sup>4</sup>

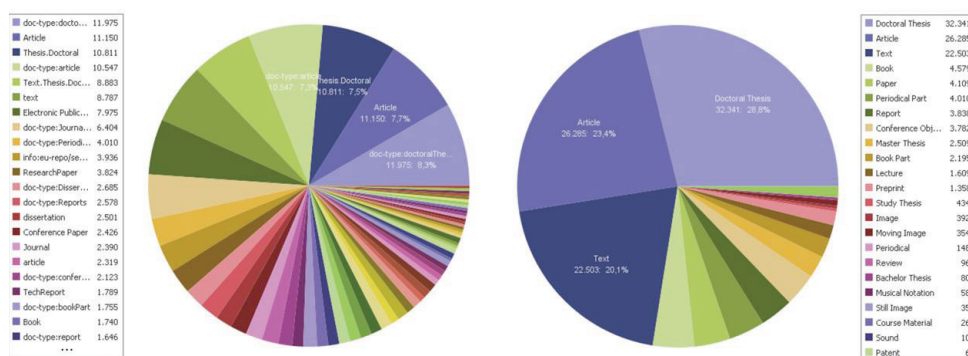


Abbildung 1: Vorher-Nachher-Vergleich der Harmonisierung von Dokumenttypen

<sup>3</sup> Stand: 15.08.2012

<sup>4</sup> Zur Unterscheidung der Dokumenttypen „Text“ und „Publication“ siehe DINI-Zertifikat 2010, S. 58.

Für den Zugriff auf den Volltext einer Informationsressource war es zum Teil ebenfalls notwendig sich diesem algorithmisch zu nähern, denn nicht immer war der direkte Link zum Volltext direkt als Identifier im Dublin Core-Metadatensatz ausgewiesen. Der Umweg führte in vielen Fällen erst über eine Metadateneinstiegsseite (Frontdoor) oder in seltenen Fällen über weitere Unterseiten zum zugehörigen Volltext.

Für wissenschaftliche Publikationen ist neben Auszeichnungselementen die fachliche Einordnung entsprechend einer Klassifikation (z. B. DDC-Klassifikation) ein wichtiges Beschreibungsmerkmal. Nur selten lassen sich fehlende semantische Informationen gänzlich automatisch generieren. Dennoch kann eine grundsätzliche fachliche Einordnung teilweise per Volltextanalyse erzeugt werden. So ließ sich eine fachliche Zuordnung über ein hash-basiertes Verfahren auf Grundlage eines vorklassifizierten Trainingsdatenraums ermitteln. Je nach Umfang des Trainingsdatenraumes für ein bestimmtes Fachgebiet war die erzeugte Information relativ zutreffend. In Fachgebieten mit hohem Trainingsdatenaufkommen (z. B. Mathematik, Physik) konnten die Dokumentmetadaten auf diese Weise ebenfalls angereichert werden.

Die Anreicherung und Angleichung von Metadaten (Dokumenttypen, Automatische Klassifikation) zwischen verschiedenen Datenquellen (Repositorien) bieten bereits eine Verbesserung der Recherchemöglichkeiten. Denn die Auffindbarkeit über Ordnungen wie Dokumenttypen oder die Fachklassifikation, die sich aus semantischen Zusatzfeldern ergeben, ermöglichen zusätzlich eine systematische Form des Auffindens von wissenschaftlicher Information.

### **Semantische Anreicherung von Informationsressourcen**

Unter der semantischen Anreicherung strukturierter Daten versteht man die explizite Klassifizierung (Klassenzuordnung), Kategorisierung mittels Ontologien und Verlinkung heterogener verteilter Uniform Resource Identifier (vgl. URI) im World Wide Web. Um referenzierbare URIs mit Algorithmen und Tools automatisch zu verlinken, müssen sie zunächst in dem vom World Wide Web Consortium (vgl. W3C) entwickelten Standardformat RDF vorhanden sein.

Die strukturierten und standardisierten RDF-Datensets können auf unterschiedlichen semantischen Ebenen miteinander verlinkt werden und bilden die Linked Data Cloud (vgl. Linked Data). Das kontinuierliche Wachstum<sup>5</sup> der in Fachkreisen so genannten "Datenwolke" lässt auf das Potential der veröffentlichten Daten hoffen, das sich der globalen Netzgemeinschaft bietet. Im Januar 2012 (vgl. Stand Datenwolke) waren ca. 31 Milliarden Informationstripel mit ca. 500 Millionen externen Links in der Datenwolke vernetzt. Projekte wie LOD2 (vgl. LOD2) und ResearchSpace (vgl. ResearchSpace) betreiben aktive Forschung auf Basis der vorhandenen Datensets.

Die Verknüpfung bzw. Verlinkung identischer relevanter Informationsressourcen im im Semantic Web ist nicht trivial. Sowohl die Identifikation potentieller identischer Ressourcen auf Klassenebene (Ontologie matching) als auch auf der Instanzebene (Metadaten) sind Forschungsthemen der Semantic Web Community. Verlinkungs- und Kontextualisierungstools wie Silk (vgl. Silk), Limes (vgl. Limes) und Cliopatria (vgl. Cliopatria) implementieren automatisierte Vergleichsverfahren für die Erzeugung neuer Relationen zwischen zwei oder mehreren RDF-strukturierten Datenpools.

Aus technischer Sicht wird somit durch offene RDF-Datenpools eine neue Vernetzung über verteilte und heterogene Datenquellen ermöglicht. Wenn alle Daten im selben offenen Standardformat vorhanden sind, können generische Applikationen entwickelt und implementiert werden. Anders als die Vernetzung und der Zugriff auf verteilte Datenbanken über REST-APIs können über verteilte RDF-Datenpools generische automatisierte Abfragen durchgeführt und Informationsressourcen aggregiert werden.

Ähnliche Ansätze werden in zahlreichen Projekten der Semantic Web Community verfolgt und teilweise erfolgreich eingesetzt. Zu erwähnen sind hier die Aktivitäten der Web-based Systems Group (vgl. Web-based Systems Group FU Berlin) der Freien Universität Berlin wie das Projekt LOD2 (vgl. LOD2) – „Creating Knowledge out of Interlinked Data“. Im Rahmen des EU geförderten Projektes DM2E – „Digitized Manuscripts to Europeana“ (vgl. DM2E) werden Annotations- und Aggregationsapplikationen (vgl. KORBO) primär für die Geisteswissenschaften auf der Basis von RDF-Daten entwickelt.

---

<sup>5</sup> Erste Abbildung des Datengraphens: [http://richard.cyganiak.de/2007/10/od/od-datasets\\_2007-05-01.png](http://richard.cyganiak.de/2007/10/od/od-datasets_2007-05-01.png),  
letzte Abbildung der Datenwolke: [http://richard.cyganiak.de/2007/10/od/od-datasets\\_2011-09-19.png](http://richard.cyganiak.de/2007/10/od/od-datasets_2011-09-19.png)

*Beispiele aus der Bibliothekswelt*

Chronologische Zusammenfassungen (Pohl, 2010 und Ostrowski, 2010) der Linked-Open-Data-Aktivitäten in der deutschen Bibliothekswelt zeigen, dass die damit verbundenen technischen Umsetzungen in der Praxis bereits vereinzelt das prototypisch-experimentelle Stadium hinter sich gelassen haben und zunehmend in die Produktivumgebungen einzelner Akteure integriert werden.

Eine der ersten öffentlichen Institutionen in Deutschland, die aktiv die Linked und Open Data-Bewegungen unterstützte und ihre Katalogdaten zunächst als Open Data (Pohl, 2010), später als Linked Data (Ostrowski & Pohl, 2010) veröffentlichte, war das Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz). Das hbz startete 2010 unter der Domäne Linked Open Bibliographic Data (vgl. Linked Open Bibliographic Data) zwei Linked Data Services: lobid-resources (vgl. lobid-resources) und lobid-organisation (vgl. lobid-organisation). Der Service lobid-resources verknüpft die internen Autoren-IDs aus dem Verbundkatalog mit den RDF-Daten des Linked Open Data Dienstes der Deutschen Nationalbibliothek (vgl. Linked Open Data Dienstes der Deutschen Nationalbibliothek). Der Service lobid-organisation definiert eindeutige Uniform Resource Identifier (vgl. URI) für alle im ISIL-Verzeichnis (vgl. ISIL-Verzeichnis) eingetragenen Institutionen und persistiert sie als referenzierbare HTTP-URIs als Knoten im RDF-Graphen.

Gleichzeitig wurden die RDF-Daten sowohl als Datendump wie auch über die SPARQL-Endpoints der lokalen Triplestores in das webbasierte Verzeichnis für RDF-Datensets eingetragen und unter der Creative Commons CC Zero-Lizenz (vgl. Creative Commons CC Zero-Lizenz) zur uneingeschränkten Weiternutzung zur Verfügung gestellt.

Neben der RDFisierung und Veröffentlichung der Katalogdaten betreiben die DNB und das HBZ kooperativ die Plattform culturegraph.org (vgl. culturegraph.org) als einen Linked Open Data Service zur Verwendung gemeinsamer Identifikatoren (URIs) für kulturelle Erzeugnisse (Bücher und andere Texte, Gemälde, Skulpturen, Musikstücke usw.).

Im Zuge dieser nationalen Aktivitäten haben weitere Institutionen und Bibliotheken ihre Kataloge und Metadaten in den RDF-Standard überführt und den Zugriff auf diese für die breite Web Community ermöglicht. Die Universitätsbibliothek Mannheim startete 2010 ein Pilot-LOD-Dienst (vgl. Pilot-LOD-Dienst der UB Mannheim) und befindet sich aktuell noch in der Entwicklungsphase.

Eine der bislang für die Rechercheoptimierung innovativsten Aktivitäten ist die Anreicherung und Vernetzung des Standard-Thesaurus Wirtschaft mit der Linked Open Data Cloud als "ein weiterer Meilenstein in der LOD-Strategie der ZBW". (zbw.eu, 2011)

Eine weitere innovative Entwicklung auf Basis semantischer Technologien ist die SLUBSemantics – Multilinguale semantische Suche (vgl. SLUBSemantics) der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB). Bei dieser werden externe RDF-Ressourcen für die automatisierte semantische Klassifizierung und Verlinkung der Suchbegriffe in Echtzeit aggregiert.

Im Rahmen der diesjährigen CeBIT2012 wurde der erste deutsche Programmierwettbewerb "Apps4Deutschland" auf der Basis von frei verfügbaren Online-Datenressourcen durchgeführt – ein wichtiger Indikator für das steigende Interesse aus der Wirtschaft für die praktisch größte Ansammlung strukturierter und freiverfügbarer Datensammlung im Netz. Den zweiten Preis in der Kategorie „Daten“ gewann B3Kat (vgl. B3Kat) mit dem Linked Open Data Service der Bayerischen Staatsbibliothek, dem Bibliotheksverbund Bayern und dem Kooperativen Bibliotheksverbund Berlin-Brandenburg. Der Service bietet Zugriff auf 23 Millionen Medien(nachweise) aus 180 wissenschaftlichen Bibliotheken in Bayern, Berlin und Brandenburg. Die Datenbasis umfasst etwa 600 Millionen RDF-Tripel oder 5,5 GB als Download.

Die kontinuierlich ansteigende Zahl der weltweit registrierten Datensets in der „Bibliographic Data Domain“ im Verzeichnis „The Data Hub“<sup>6</sup> (vgl. The Data Hub) bietet einerseits vielfältige Möglichkeiten zur Verknüpfung mit Informationsressourcen der Repositorien und verdeutlicht andererseits den Trend der Veröffentlichung bibliografischer Daten bzw. der Metadaten sowie der konkreten Publikationen selbst. Die Weiterentwicklung und Implementierung von innovativen Services für Repositorien in Deutschland ist in diesem Zusammenhang ist notwendig.

---

<sup>6</sup> Mit dem Stand von Juli 2012 sind es 88 Datensets.

*Semantische Anreicherung von Metadaten am Beispiel von OA-Netzwerk*

Der folgende Abschnitt erläutert und dokumentiert die Aktivitäten aus dem OA-Netzwerk-Arbeitspaket mORE (modelling Object Reuse and Exchange). Ziel dieses Arbeitspakets war die Evaluierung des Datenraums von OA-Netzwerk mit Blick auf die Rolle eines potentiellen Aggregators für Open-Access-Publikationen sowohl im Kontext der Deutschen Digitalen Bibliothek (vgl. Deutsche Digitale Bibliothek) als auch international als Datenlieferant für Europeana (vgl. Europeana) bzw. EuropeanaResearch (vgl. Europeana Research).

Die technische Grundlage für die semantische Anreicherung des OA-Netzwerk-Datenraums bilden W3C-Standards wie RDF, RDFS (vgl. RDF-Schema), OWL (vgl. Web Ontology Language) sowie der Standard der Open Archives Initiative (vgl. OAI) Object Reuse and Exchange (vgl. OAI-ORE) für die eindeutige Identifizierung verteilter Web-Ressourcen.

Die über OAI-PMH geharvesteten Metadaten basieren auf dem relativ flachen Dublin Core-Datenmodell. Aufgrund dieser im semantischen Sinne eingeschränkten Ausgangs-Datenbasis war es nicht möglich, tiefere semantische Beziehungen über die durch das Harvesting gesammelten Metadaten direkt zu identifizieren. Demzufolge lag der Fokus der Arbeiten auf der Rolle von OA-Netzwerk als Aggregationsknoten, der Kontextualisierung und externen Verknüpfung der im Datenraum vorhandenen Metadaten, sowie der Bereitstellung des angereicherten Datenraums als Linked Open Data.

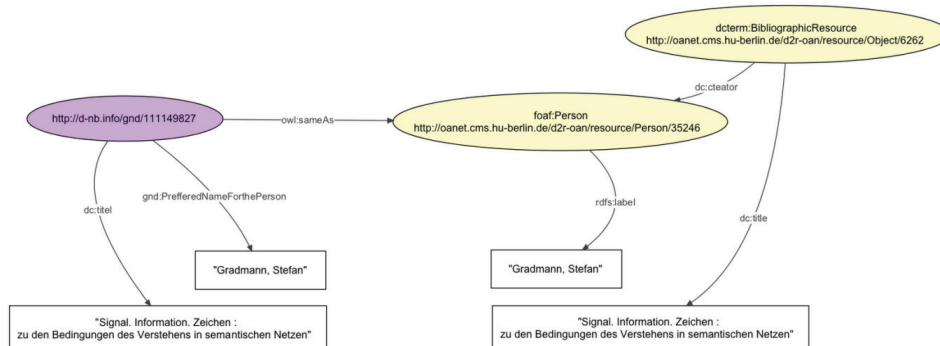
Um die in einem relationalen Datenbanksystem gespeicherten Metadaten in RDF zu überführen, haben wir nach einer Lösung gesucht, die nicht auf Austauschformaten beruht, sondern den Zugriff auf alle Relationen und Metadaten des Systems zu erzeugen. Für die Konvertierung haben wir auf Open-Source-Tools und daher die Lösungen der Web-based System Group (siehe oben) integriert und an das vorhandene System angepasst (vgl. Iwanowa, 2011).

### *Datenaufbereitung und Strategien der externen Verlinkung*

Für die Überführung der DC-Metadaten von OA-Netzwerk nach RDF wurde der D2R-Server (vgl. D2R-Server) an die vorhandene Datenbank angepasst und in das Backend von OA-Netzwerk integriert. Um eine Datenintegration und Kontextualisierung zwischen Europeana und OA-Netzwerk zu erreichen, wurde zudem das Datenmodell von Europeana (vgl. Europeana Datenmodell) eingebunden. Der D2R-Server ermöglicht ein on-the-fly-Mapping der Metadaten auf beliebige RDF-basierte Zielformate und bietet zugleich eine SPARQL-Schnittstelle zur Abfrage des Datenpools. Zudem kann mithilfe des D2R-Tools ein RDF-Dump der gesamten Datenbasis erzeugt werden. Letztere Variante wird aus Performanzgründen vorgezogen, um bspw. die RDF-Triples weiterzuverarbeiten oder mit externen Quellen zu verknüpfen.

Nach der Konvertierung der Metadaten nach RDF wurde die Verknüpfung der Personennamen von OA-Netzwerk mit den RDF-basierten Personendaten aus der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek (vgl. Dokumentation des Linked Data Services der DNB) prototypisch umgesetzt. Dafür wurden die beiden Open-Source-Tools LIMES (vgl. LIMES) und SILK (vgl. SILK) getestet und angewandt. Die Tools können auf Basis von String-Matching-Algorithmen Verknüpfungen zwischen verschiedenen Datenquellen unter Verwendung von semantisch vergleichbaren Feldern erzeugen. Beide Tools erlauben dabei sowohl die Einstellung für eine exakte Übereinstimmung der zu vergleichenden Literale als auch eine flexible Ähnlichkeitsangabe. Allein bei der Umsetzung des ersten Falls zeigte sich, dass 40.000 Personennamenlitterale aus OA-Netzwerk und der DNB diese exakte Übereinstimmung erfüllten. Für die Identifizierung und Verlinkung der Personennamen beider Datenquellen wurde eine einfache Heuristik entwickelt. Eine genaue Analyse der beiden RDF-Datenpools zeigte, dass für die Disambiguierung von Personennamen neben dem Titel einer Publikation nur wenige andere Datenfelder herangezogen werden können (Institution des publizierten Werkes, Publikationsjahr). Mit der Erweiterung der RDF-Daten der DNB um die Titeldaten ist es nun maschinell möglich, neben Personennamen auch Titeldaten zu vergleichen. Mit den vorhandenen Matching-Algorithmen ist es beim gleichzeitigen Vergleich der o. g. Metadaten möglich, Autoren mit einer sehr hohen

Wahrscheinlichkeit (>99%) zu identifizieren und eine *owl:sameAs*-Beziehung zwischen den beiden Datenräumen zu erstellen. Abbildung 2 visualisiert beispielhaft die Beziehungen der beiden Datenbasen.



**Abbildung 2: Autorentifizierung an Beispiel von OAN und DNB**

Das Beispiel zeigt den Mehrwert, der aus scheinbar zusammenhangslosen Daten generiert werden kann. Über die Identifikation von Informationsressourcen ist eine Einordnung in den globalen Kontext anderer Datenräume möglich. Zusatzinformationen lassen sich somit über diese Identifikation dem globalen Kontext einer identifizierten Ressource hinzufügen.

Die Entwicklung des Semantic Web kann nur dann dauerhaft erfolgreich und nachhaltig fortgeführt werden, wenn automatisierte Verlinkungsverfahren entwickelt werden, die in der Lage sind, die stark ansteigende Menge an strukturierten Daten zu verarbeiten. WissenschaftlerInnen ermöglicht dies, sowohl AutorInnenidentitäten als auch jegliche Arten digitaler Objekte im Netz bereits während der Produktionsphase oder im Publikationsprozess miteinander zu verlinken.

Der in der zweiten Projektphase entwickelte Linked Open Data-Dienst von OA-Netzwerk ist ein experimenteller Betaservice mit Entwicklungspotenzial. Speziell für Repositorien ist ein vollautomatisierter Dienst zur externen Entitäten-Identifikation und entsprechenden Verlinkung eine sinnvolle Weiterentwicklung.

Als Ergebnis des Arbeitspakets wurde die angereicherte OA-Netzwerk-Datenbasis als Linked Open Data veröffentlicht (vgl. OA-Netzwerk als Linked Open Data) und für



externe Nutzer zur Verfügung gestellt (OA-Netzwerk bei The Data Hub). Die Daten sind frei verfügbar und unter CC0-Lizenz ohne Einschränkung nutzbar. Die lizenzfreie Datenveröffentlichung wird einerseits die Linked Open Data Community und andererseits die Entwicklung und Implementierung neuer Software auf dieser Datenbasis unterstützen.

### **Diskussion und Ausblick**

Die Ergebnisse von OA-Netzwerk stellen einen wichtigen Schritt zu einer verbesserten Informationsversorgung im Open-Access-Bereich dar. Sie können als technologische Grundlage für bereits entwickelte oder noch in Planung befindliche Mehrwertdienste dienen. Denkbar ist u. a. eine Verbesserung der Sichtbarkeit für Datenaggregatoren und Anbieter übergreifender und auf den bereitgestellten Daten basierender Dienste im nationalen und internationalen Kontext. Für Repositorien lassen sich derartige Effekte dank der Einbindung ihrer Bestände in Informationsnetzwerke wie OA-Netzwerk feststellen. Mittelbar unterstützt OA-Netzwerk Wissenschaftler folglich sowohl bei der Informationsrecherche als auch bei der Publikation eigener Ergebnisse. Zusätzlich ermöglicht die Analyse der über OA-Netzwerk und die angeschlossenen Dienste verfügbaren Daten Rückschlüsse zu den bis dato realisierten Publikationen in deutschen Repositorien, zur Art und Weise der Erschließung sowie zu Herausforderungen an diese im Hinblick auf die Integration des wissenschaftlichen Publikationsoutputs im Semantic Web.

Das Projekt OA-Netzwerk ist zum Zeitpunkt der Präsentation im Rahmen der WissKom 2012 bereits beendet. Zur Verstetigung der Projektergebnisse wurden verschiedene Szenarien zur Integration in die nationale und internationale Informationsinfrastrukturlandschaft evaluiert. Eine Vernetzung innerhalb der Wissenschafts- und Infrastrukturcommunity in organisatorischer Hinsicht scheint ebenso erforderlich wie etwa die langfristig stabile Verknüpfung und die semantische Anreicherung der Daten auf einer technischen Ebene. Über die beschriebenen Formen der Informationsaufbereitung werden der Mehrwert vernetzter Systeme und gleichzeitig die Bedeutung des Überwindens isolierter Datenquellen hin zu einer größtmöglichen Datenverflechtung über lokale Systemgrenzen hinweg deutlich.

Die Aufwertung mit Zusatzinformationen und direkt verlinkten zugehörigen Ressourcen wie beispielsweise referenzierten Dokumenten ermöglicht eine effizientere Recherche. Zusätzlich werden durch Serendipitätseffekte, also dem zufälligen Auffinden interessanter Informationen, neue Türen neben der klassischen Recherche geöffnet.

Mit Abschluss des Projekts OA-Netzwerk wurden sämtliche im Projekt vorgesehen Ziele erreicht. Die entstandene Infrastruktur umfasst diverse modulare Dienste, die in ein nahezu vollautomatisches System integriert sind. Weitere Repositorien lassen sich mit wenigen Klicks in den OAN-Workflow aufnehmen. Zudem bietet die Plattform viele Möglichkeiten für Weiterentwicklungen.

Ein wichtiger Aspekt in der Weiterführung von OA-Netzwerk ist die bereits erwähnte Erweiterung des Datenraums über DINI-zertifizierte Repositorien hinaus. Sie wird zum Zeitpunkt des Vortrags abgeschlossen sein und positioniert OA-Netzwerk ein Stück weit mehr als nationalen Aggregator und Datenlieferant für übernationale Dienste- und Datenanbieter.

Auch bleibt zu beobachten wie sich die Repositorienlandschaft in Bezug auf mehrteilige oder komplexe Ressourcen verhält (Enhanced Publications). Bislang sind Informationen über die Beziehung zusammengehörender Ressourcen selten. Wird er doch ausgewiesen, dann oft in nicht standardisierter Form. Spezifikationen für die Erweiterung eigener OAI-PMH Schnittstellen sind in Form von OAI-ORE bereits seit mehreren Jahren verfügbar. Der Mangel an geeigneten Tools bzw. die Unterstützung für vorhandene Repositoriensysteme erschwert die Aufnahme komplexer Ressourcen. Im DFG-Projekt eco4r, einer Partnerschaft des HBZ Köln und der Universitätsbibliothek Bielefeld, wurden zu diesem Zweck Plugins zur Abbildung mehrteiliger Publikationen (Compound Objects) für die Repositoriensysteme Fedora und OPUS entwickelt (vgl. ECO4R). Sollte sich ein deutlicher Trend in diese Richtung abzeichnen, ist die Unterstützung mehrteiliger Ressourcen auch für OA-Netzwerk perspektivisch eine geeignete Erweiterung.

Im Bereich der semantischen Vernetzung ist die Unterstützung eines Autoren-identifikationsdienstes naheliegend, um die im Datenraum enthaltenen Autoren sukzessive identifizieren und auf Basis dieses Ankers mit anderen Datenquellen zu

verknüpfen. Die Entwicklungen in dem zunehmend für notwendig angesehenen Feld der Autorenidentifizierung sind zu beobachten. Es bleibt abzuwarten, ob sich Konzepte von Organisationen wie ORCID zu einem praktischen Standard entwickeln.

Mit Blick auf eine Steigerung der Sichtbarkeit der in Repositoriensystemen enthaltenen Informationsobjekte, wäre eine Anbindung an Systeme außerhalb des eigenen Informationssystems denkbar, z. B. die Kopplung an bestehende Bibliothekssysteme. Beide Informationswelten könnten durch eine Zusammenführung profitieren. Ein Aggregatordienst wie OA-Netzwerk könnte dabei sowohl die Funktion der Bereinigung (Harmonisierung, Anreicherung) der repositorienseitigen Metadaten übernehmen, als auch als zentralisierte Schnittstelle fungieren, um eine Anbindung einzelner kleiner Repositorien zu vermeiden.

#### **Literaturverzeichnis**

Die Webquellen wurden zuletzt am 16. August 2012 aufgerufen.

Aschenbrenner, Andreas; Blanke, Tobias; Küster, Marc W.; Pempe, Wolfgang (2010). Towards an Open Repository Environment. In: Journal of Digital Information, Vol 11, No 1 (2010). <http://journals.tdl.org/jodi/article/view/758>

B3Kat = <http://lod.b3kat.de/doc>

Creative Commons CC Zero-Lizenz = <http://opendefinition.org/licenses/cc-zero/>

Cliopatria = <http://cliopatria.swi-prolog.org/home>

culturegraph.org = <http://www.culturegraph.org/>

D2R-Server = <http://d2rq.org/d2r-server>

Deutsche Digitale Bibliothek = <http://www.deutsche-digitale-bibliothek.de/>

DINI-Liste der Repositorien = <http://www.dini.de/dini-zertifikat/liste-der-repositorien/>

DINI Zertifikat 2010 = <http://nbn-resolving.de/urn:nbn:de:kobv:11-100182794>

DM2E = <http://dm2e.eu/>

Dokumentation des Linked Data Services der DNB = <https://wiki.d-nb.de/display/LDS/Dokumentation+des+Linked+Data+Services+der+DNB?focusedCommentId=57901372>

eco4r = <http://www.eco4r.org/>

Europeana = <http://www.europeana.eu/portal/>

Europeana Research = <http://www.europeana-libraries.eu/documents/868553/bc8a98bc-5339-4117-bb2b-4cb5f75b7aaf>

Europeana Datenmodell = <http://pro.europeana.eu/web/guest/edm-documentation>

Glaser, Hugh; Halpin, Harry: The linked data strategy for global identity. *IEEE Internet Computing*, 2012, 16, (2), 68-71. (doi:10.1109/MIC.2012.39)

Halpin, Harry et al.: When owl:sameas Isn't the Same: An Analysis of Identity in Linked Data. Proc. 9th Int'l Semantic Web Conf. (ISWC 10), Springer, 2010, pp. 305-320; <http://data.semanticweb.org/conference/iswc/2010/paper/261>

Heath, Tom; Bizer, Christian: *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers, ISBN: 978160845431, 2011. Free HTML version: <http://linkeddatabook.com/>

Heath, Tom: Linked Data — Welcome to the Data Network, *IEEE Internet Computing*, vol. 15, no. 6, 2011, pp. 70-73; <http://dx.doi.org/10.1109/MIC.2011.153>.

ISIL-Verzeichnis = <http://dispatch.opac.d-nb.de/DB=1.1/DB=1.2/>

Iwanowa, Julia: Kurze Einführung in Semantic Web Standards. In: OA-Netzwerk Blog am 5.12.11 = <http://oanetwerk.wordpress.com/2011/12/05/oan-lod-dienste-teil-1/>

KORBO = <http://korbo.muruca.org/>, <http://thepund.it/>

LIMES = <http://aksw.org/Projects/limes>

Linked Data = <http://linkeddata.org/>

Linked Open Bibliographic Data = <http://lobid.org>

Linked Open Data Dienstes der Deutschen Nationalbibliothek = [http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata\\_node.html](http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata_node.html)

lobid-organisation = <http://lobid.org/organisation>

lobid-resources = <http://lobid.org/resource>

LOD2 = <http://lod2.eu/Welcome.html>

Müller, Thomas; Severiens, Thomas; Malitz, Robin; Schirmbacher, Peter: OA Network. An Integrative Open Access Infrastructure in Germany. In: *D-Lib Magazine* (15) 9/10, 2009. = <http://www.dlib.org/dlib/september09/mueller/09mueller.html>

Nikolov, Andriy and d'Aquin, Mathieu: Identifying relevant sources for data linking using a semantic web index. In: WWW2011 Workshop: Linked Data on the Web (LDOW 2011) at 20th International World Wide web Conference (WWW 2011), 29 March 2011, Hyderabad, India.

OAN-Suche = <http://oansuche.open-access.net/>

OA-Netzwerk als Linked Open Data = <http://oanet.cms.hu-berlin.de/d2r-oan/>

OA-Netzwerk bei The Data Hub = <http://thedatahub.org/dataset/oanetwerk>

OAI = <http://www.openarchives.org/>

OAI-ORE = <http://www.openarchives.org/ore/1.0/primer.html>

OpenDOAR = <http://www.opendoar.org/find.php?format=charts>

Ostrowski, Felix; Pohl, Adrian: 'Linked Data' - und warum wir uns im hbz-Verbund damit beschäftigen. Preprint. Erschienen in B.I.T. Online, 13/3, S. 259-268)

Pilot-LOD-Dienst der UB Mannheim = <http://data.bib.uni-mannheim.de/>

Pohl, Adrian (2010): Open Data im hbz-Verbund. Preprint. Erschienen in ProLibris, 3/2010, S. 109-113)

RDF = <http://www.w3.org/RDF/>

RDF-Schema = <http://www.w3.org/TR/rdf-schema/>

ResearchSpace = <http://www.researchspace.org/>

SILK = <http://www4.wiwiss.fu-berlin.de/bizer/silk/>

SLUBSemantics = <http://www.slub-dresden.de/ueber-uns/projekte/infrastruktur-und-softwareentwicklung/slubsemantics/>

Stand Datenwolke = <http://www4.wiwiss.fu-berlin.de/lodcloud/state/>

The Data Hub = <http://thedatahub.org/group/bibliographic?page=3>

URI = <http://tools.ietf.org/html/rfc3986>, <http://www.ietf.org/rfc/rfc3986.txt>,  
[http://de.wikipedia.org/wiki/Uniform Resource Identifier](http://de.wikipedia.org/wiki/Uniform_Resource_Identifier)

W3C = <http://www.w3.org/>

Web Ontology Language = <http://www.w3.org/TR/owl-features/>

Web-based Systems Group = <http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/>

zbw.eu, 2011: [http://www.zbw.eu/presse/pressemitteilungen/2011\\_11\\_17.htm](http://www.zbw.eu/presse/pressemitteilungen/2011_11_17.htm)

## **Kooperation und Vernetzung als Kernaufgabe disziplinspezifischer Versorgung mit Volltexten: Das Fachrepositorium „OstDok“**

Doris Škarić

Bayerische Staatsbibliothek, München

### **Zusammenfassung**

Das DFG- geförderte Fachrepositorium „OstDok – Osteuropa-Dokumente online“ ist ein Kooperationsprojekt, in dem sich bibliothekarische, informationswissenschaftliche und fachwissenschaftliche Kompetenzen ergänzen und dadurch entstehende Synergien wechselseitig genutzt werden. Neben den Trägerinstitutionen bestehen enge Kooperationen mit anderen Instituten der Osteuropaforschung, die hauptsächlich dem weiteren inhaltlichen Ausbau dienen. Durch die Vernetzung mit anderen Repositorien, Plattformen und weiteren Online-Angeboten werden der Bekanntheitsgrad und die Sichtbarkeit von OstDok weiter gesteigert.

### **Abstract**

The DFG-funded repository „OstDok – Osteuropa-Dokumente online“ is a cooperative project where the expertise of library science, information science and of scientific experts complements each other and leads to synergy. Besides the responsible institution there is close cooperation with other institutions in the field of East European Studies, which mainly serve the increase of content. The level of awareness and visibility of OstDok will be raised by networking with other repositories, platforms and web offers.

Das Fachrepositorium „OstDok“ („Osteuropa-Dokumente online“) ist die zentrale Recherche- und Publikationsplattform im Internet für frei zugängliche Volltexte der Osteuropastudien<sup>1</sup>. Auf OstDok werden die Interessen verschiedener Akteure rund um das elektronische Dokument gebündelt: der Zugriff durch Leser, das Erschließen, Bewahren und Bereitstellen durch Bibliothekare sowie das Publizieren von fachrelevanten Inhalten durch Einzelautoren, Institutionen, Projekte und Verlage.

---

<sup>1</sup> [www.ostdok.de](http://www.ostdok.de). Auf diesen und alle im Folgenden angeführten Links wurde am 15.08.2012 zugegriffen.

Der Bezugsraum des Portals erstreckt sich über Ostmitteleuropa, Südosteuropa und das östliche Osteuropa einschließlich historischer oder kultureller Bezugsräume wie etwa die früheren Sowjetrepubliken oder aber Finnland (bis 1918). Typische Materialien auf OstDok sind durchsuchbare Texte, bei denen es sich um Retrodigitalisate und genuin oder parallel elektronisch vorhandene Dokumente handelt<sup>2</sup>. Ihre Einbindung in die Plattform ist durchweg rechtlich zulässig<sup>3</sup>. Bevorzugte Dokumentsprachen sind westliche oder in Osteuropa vorherrschende Sprachen in den jeweiligen Schriftsystemen<sup>4</sup>.

OstDok bildet ein gemeinsames „Dach“ für unterschiedliche Disziplinen wie etwa Osteuropäische Geschichte, Slavistik oder aber osteuropabezogene Musikwissenschaft, Ethnologie, Theologie, Politik u.a.<sup>5</sup>. Durch seine weltweit uneingeschränkte Verfügbarkeit wird es zudem helfen, die unzweifelhaft vorhandenen Barrieren in der jeweiligen Rezeption zwischen osteuropäischer und westlicher Forschung zu überwinden – auch wenn sich der Fokus des von der Deutschen Forschungsgemeinschaft geförderten Projekts vorwiegend auf die Osteuropa-Forschung des deutschsprachigen Raums richtet. Die Kooperations- und Vernetzungsstrategie bildet dabei eine tragende Säule beim Auf- und Ausbau der Plattform. Dies verdeutlicht schon die Trägerstruktur von OstDok.

---

<sup>2</sup> Künftig sollen zudem weitere Ressourcentypen wie Forschungsdaten oder Metadaten relevanter Ressourcen auf qualitätsgesicherten externen Servern eingebunden werden.

<sup>3</sup> Die Materialien sind entweder gemeinfrei oder können im Ergebnis einer Nutzungsrechteinräumung durch den jeweiligen Inhaber des Verwertungsrechts bereitgestellt werden.

<sup>4</sup> Ausführlicher zum Profil siehe Norbert Kunz, Dokumente aller Länder vereinigt Euch! – das Fachrepositorium „Osteuropa-Dokumente Online“ (OstDok), in: „Ost- und Südosteuropakompetenz – Voraussetzung für erfolgreiche Integration“, 40. ABDOS-Tagung, Referate und Beiträge, zsgest. von Franz Görner, Berlin 2012, S. 24-29 (Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Veröffentlichungen der Osteuropa-Abteilung, Bd. 40).

<sup>5</sup> Aktuell weisen über 50% aller enthaltenen Publikationen einen Bezug auf zur Osteuropäischen Geschichte, 42% zur Politik und 28% zur Philologie auf (Stand: August 2012). Weitere Schwerpunkte bilden die Bereiche Wirtschaft, Gesellschaft / Volkskunde / Brauchtum, Militär, Geographie / Heimat-/ Länderkunde / Reisen, Recht/ Verfassung / Verwaltung, Kirche / Religion, Philosophie, Bildung / Wissenschaft, Bibliotheks- und Informationswissenschaft, Journalismus / Publizistik / Medien, Kultur allgemein / Kunst/ Architektur, Musik / Theater/ Tanz, Naturwissenschaft / Technik / Medizin / Psychologie / Sport.

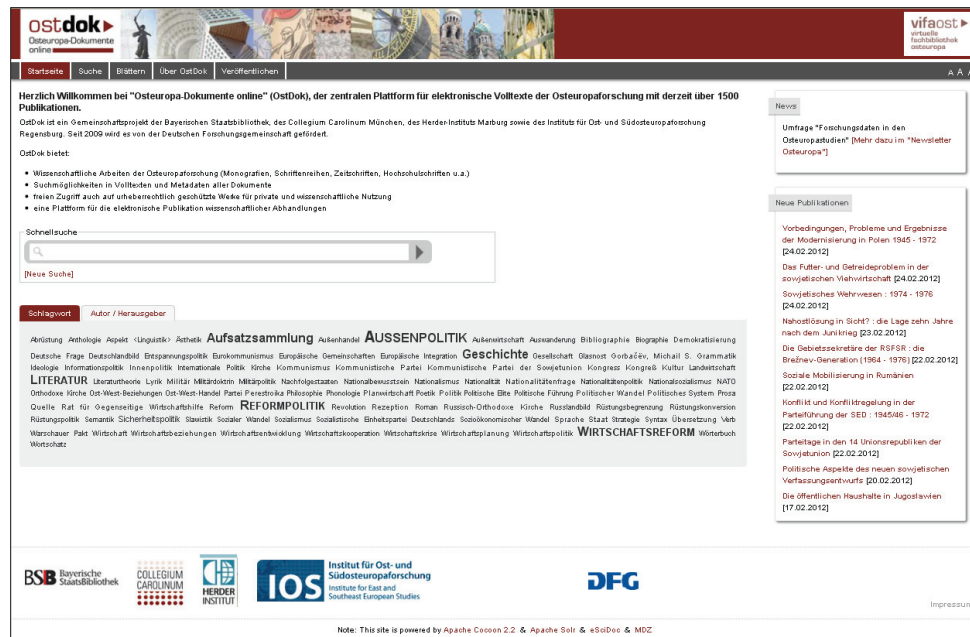


Abbildung 1 Startseite der OstDok-Plattform

## Träger und Beiräte: Synergie durch Vielfalt

OstDok ist ausdrücklich ein Kooperationsprojekt, in dem sich bibliothekarische, informationswissenschaftliche und fachwissenschaftliche Kompetenzen ergänzen und Synergien wechselseitig genutzt werden. Schon unter den Projektträgern finden sich daher neben einer Bibliothek auch fachwissenschaftliche Einrichtungen, wodurch eine enge Anbindung an die Fachwelt erzielt wird<sup>6</sup>.

Gesamtkoordination, technischer Systemaufbau und –betrieb sowie die Langzeitarchivierung sämtlicher Materialien obliegen bei OstDok der Bayerischen Staatsbibliothek München (BSB). Sie beherbergt eine der weltweit bedeutendsten Osteuropasammlungen (allein mehr als 1 Mio. gedruckte Bände) und betreut mehrere für die Osteuropaforschung einschlägige Sondersammelgebiete (SSGs)<sup>7</sup>,

<sup>6</sup> Die OstDok-Konzeption folgt hier weitgehend der des Schwesterportals „Virtuelle Fachbibliothek Osteuropa“ (ViFaOst, [www.vifaost.de](http://www.vifaost.de)), da in beiden Portalen die Trägerstruktur nahezu identisch ist. Die ViFaOst ist das zentrale Regionalportal für Fachinformationen zu den Osteuropastudien.

<sup>7</sup> Hierzu zählen etwa die SSGs „Ost-, Ostmittel- und Südosteuropa“ (7.41), „Rumänische Sprache und Literatur“ (7.38), „Albanische Sprache und Literatur“ (7.43), „Neuzeitliches Griechenland“ (6.16). Zur Osteuropasammlung siehe <http://www.bsb-muenchen.de/Osteuropaabteilung.286.0.html>.



weshalb ihr in der überregionalen Literaturversorgung zu Ost-, Ostmittel-, und Südosteuropa eine besondere Stellung zukommt. Mit dem Münchner Digitalisierungszentrum (MDZ) verfügt die BSB zudem über langjährige herausragende Kompetenz in den Bereichen Digitalisierung und Langzeitarchivierung<sup>8</sup>.

Die stärker fachwissenschaftlich ausgerichteten Träger – das Collegium Carolinum München<sup>9</sup> (CC), das Herder-Institut Marburg<sup>10</sup> (HI) und das Institut für Ost- und Südosteuropaforschung Regensburg (IOS)<sup>11</sup> – verfügen ebenfalls über jeweils eigene bibliothekarische Einrichtungen, sind jedoch in erster Linie im Bereich der Inhaltsakquise tätig. Ihre Aufgaben erfüllen sie dabei im Einklang mit der jeweiligen Forschungsausrichtung der Institution. Liegt der geographische Schwerpunkt des CC als Forschungsstelle für die böhmischen Länder in OstDok auf Publikationen aus Tschechien und der Slowakei, so zeichnet das HI als zentrale Forschungsstätte und wissenschaftliche Serviceeinrichtung der historischen Ostmitteleuropa-Forschung im Projekt verantwortlich für Materialien des restlichen ostmitteleuropäischen Raums. Das IOS, Anfang 2012 aus der Fusion von Osteuropa-Institut Regensburg und Südost-Institut Regensburg hervorgegangen, bündelt dagegen geschichts- und wirtschaftswissenschaftliche Forschungskompetenz zum östlichen und südöstlichen Europa, die entsprechend in das Projekt eingebracht wird.

Alle Partner geben Reihen und/oder Zeitschriften heraus, die sie als institutionelle Veröffentlichungen in OstDok einbringen. Dies geschieht bei älteren Titeln in Form von Retrodigitalisaten, bei neueren als genuin elektronische Publikationen („born digitals“). Durch die institutsinternen Kontaktstrukturen sorgen sie zudem für die Einwerbung einfacher Nutzungsrechte bei qualitativ hochwertigen Arbeiten einzelner Autoren. Das Collegium Carolinum etwa betreut in OstDok zusätzlich eine eigens gegründete Reihe herausragender Graduierungsschriften mit einer deutschen und einer tschechischen Unterreihe (letztere in enger Kooperation mit tschechischen Universitäten). In diesem Zuge wird OstDok auch über die deutschen Grenzen

---

<sup>8</sup> Siehe <http://www.digitale-sammlungen.de/>, <http://www.babs-muenchen.de/>. Aktuell stellt das Münchener Digitalisierungszentrum rund 840.000 digitalisierte Titel bereit (Stand: 8/2012). Mittelfristig wird an der BSB der gesamte gemeinfreie Bestand (ca. 100.000 Werke der Osteuropasammlung) im Rahmen einer Public-Private-Partnership mit google digitalisiert und bereitgestellt werden

<sup>9</sup> [www.collegium-carolinum.de/](http://www.collegium-carolinum.de/)

<sup>10</sup> [www.herder-institut.de/](http://www.herder-institut.de/)

<sup>11</sup> <http://www.ios-regensburg.de>

hinaus bekannt gemacht. Der direkte Kontakt, den die fachwissenschaftlichen Träger in die jeweiligen Fachkreise unterhalten, ist dabei für Akzeptanz und Bekanntheit der Plattform und ihrer Angebote von zentraler Bedeutung. Letztlich propagieren alle Trägerinstitutionen durch Fachpublikationen und bei eigenen und externen Tagungen, Konferenzen etc. die Projekthalte und bemühen sich in ihren jeweiligen Kompetenzbereichen um weitere Kooperationspartner.

Ähnlich der Trägergemeinschaft setzt sich auch der sechsköpfige OstDok-Beirat gezielt aus Vertretern verschiedener wissenschaftlicher Wirkungskreise – darunter die Bereiche „Osteuropastudien“, „Geschichte“, „Elektronisches Publizieren“ und „Informationswesen“ – zusammen<sup>12</sup>. Der Beirat berät die Träger hinsichtlich der konzeptionellen Ausrichtung der Plattform. Gerade durch die interdisziplinäre Zusammensetzung des Gremiums erhält die konzeptionelle OstDok-Ausrichtung z.B. hinsichtlich der Nutzerbedarfe oder innovativer Informationsstrukturen dabei eine vielschichtige fruchtbare Expertise.

### **Kooperation**

Da sich Fachrepositorien gerade durch ihren überinstitutionellen, auf die gesamte Fachwissenschaft ausgerichteten Charakter auszeichnen, liegt es auf der Hand, dass auch in OstDok der Zusammenarbeit mit externen Partnern höchste Bedeutung zukommt. Für alle beteiligten Akteure müssen dabei entsprechende Anreize geschaffen werden, damit sich die Kooperation beiderseitig als sinnvoll erweist.

Mit Blick auf die Akzeptanz der Plattform durch ihre Nutzer ist es unerlässlich, dass die enthaltenen Materialien a) einschlägig und qualitätsgesichert sind und b) in umfangreichem Maße zur Verfügung stehen. Zielsetzung von OstDok ist es daher, in größtmöglichem Umfang einschlägige Partnereinrichtungen zu gewinnen, die eigene (und somit qua institutionem qualitätsgesicherte) Materialien einbringen. Die Kooperationspartner können im Gegenzug die im Projekt geschaffene Infrastruktur für Digitalisierung, Volltextaufbereitung, Verwaltung, Bereitstellung und Langzeitarchivierung der Materialien nutzen und sich zudem in institutionellen

---

<sup>12</sup> Beiratsmitglieder sind: Prof. Helmut Altrichter (Erlangen), Prof. Christoph Boyer (Salzburg), Prof. Gudrun Gersmann (Paris), Wolfgang Klotz (Belgrad), Jan Plamper, Ph.D. (Berlin), Prof. Marc Rittberger (Frankfurt/Main).

Sichten im Kontext des eigenen Dokumentenbestands auf der Plattform präsentieren. Dass diese Option weithin als überaus sinnvoll erachtet wird, belegt, dass dem Projekt schon jetzt über 20 entsprechende Kooperationszusagen von fachwissenschaftlichen oder bibliothekarischen Einrichtungen, Portalen, Projekten oder Verlagen vorliegen. Hierzu zählen, um nur einige zu nennen, die Bibliothek der Friedrich-Ebert-Stiftung in Bonn<sup>13</sup> (historische und politikwissenschaftliche Quellen), die Martin-Opitz-Bibliothek Herne<sup>14</sup> (historische Quellen), die Schweizerische Osteuropabibliothek Bern<sup>15</sup> (Schweizerische Osteuropa-Monografien), die Universitätsbibliothek Greifswald<sup>16</sup> (Dokumente zum Sondersammelgebiet Baltische Länder), die Plattform russland.RU<sup>17</sup> (Analysen, Berichte etc.) sowie der Verlag Otto Sagner<sup>18</sup> (Osteuropa-Monografien). Ein fruchtbares Feld sind zudem „indirekte“ Kooperationen, die sich durch die Nachnutzung von in anderen Projekten entstandenen Digitalisaten ergeben, an denen einer der Träger beteiligt ist. So konnten beispielsweise über das Projekt Digi20 aus namhaften Verlagsprogrammen bereits rund 700 teils hochaktuelle Publikationen zu einer Osteuropa-Thematik bereitgestellt werden (u.a. Verlage Otto Sagner, Schöningh, Vandenhoeck & Ruprecht<sup>19</sup>).

Bei allen Digitalisierungsaktivitäten soll in OstDok der viel kritisierten Praxis der Paralleldigitalisierung entgegengewirkt werden. Bereits frei zugängliche Publikationen werden grundsätzlich nicht zweigitalisiert. Vielmehr sollen im Rahmen von Absprachen einvernehmliche Arbeitsteilungen bezüglich Digitalisierung, Bereitstellung, Langzeitarchivierung verabredet werden. Hervorzuheben ist hier die mit der Stiftung Wissenschaft und Politik in Berlin (SWP) getroffene Regelung<sup>20</sup>.

---

<sup>13</sup> <http://library.fes.de/>.

<sup>14</sup> <http://www.martin-opitz-bibliothek.de/>.

<sup>15</sup> <http://www.ub.unibe.ch/sob/content/>.

<sup>16</sup> <http://www.uni-greifswald.de/bibliothek.html>.

<sup>17</sup> <http://www.russland.ru/>.

<sup>18</sup> <http://verlag.kubon-sagner.de/>.

<sup>19</sup> <http://digi20.digitale-sammlungen.de/>. Des weiteren nutzt OstDok Inhalte nach, die in den Projekten Leibniz publik (<http://www.leibniz-publik.de/>), entstanden sind.

<sup>20</sup> <http://www.swp-berlin.org>.

In Absprache mit der Rechtsnachfolgerin des Bundesinstituts für Ostwissenschaftliche Studien<sup>21</sup> wurden im Rahmen von OstDok die älteren „Berichte des Bundesinstituts [...]“ (bis 1994) digitalisiert und bereitgestellt, während die jüngeren Jahrgänge (ab 1995, Verlags-PDF) in der sozialwissenschaftlichen Plattform SSOAR<sup>22</sup> angeboten werden. In einem zweiten Schritt soll dann auf den jeweils komplementären Bestand der anderen Plattform verlinkt werden<sup>23</sup>. Derartige Kooperationen gilt es künftig zu intensivieren, wobei mithin bereits bestehende Kontakte wie etwa die Partnerschaft zwischen der Bayerischen Staatsbibliothek mit der neu gegründeten Russischen Präsidentenbibliothek im Bereich des Datenaustauschs nur mehr konkretisiert werden müssen<sup>24</sup>.

Spezielle Anreize zur Zusammenarbeit sollen künftig für außeruniversitäre wissenschaftliche Osteuropa-Einrichtungen des deutschsprachigen Raums geschaffen werden. Sie sollen künftig verstärkt dafür gewonnen werden, neu entstehende eigene Publikationen direkt genuin oder parallel elektronisch im open access auf OstDok zu publizieren. Es bietet sich an, dazu neue Reihen zu gründen und auf OstDok prominent zu präsentieren. Daneben können aber auch bereits im Druck vorliegende institutseigene Publikationen im Rahmen des OstDok-Projekts retrodigitalisiert, langzeitarchiviert und bereitgestellt werden. Eine weitere Form der Zusammenarbeit zielt darauf ab, den Einrichtungen den Zugriff auf wichtige forschungsrelevante Literatur erheblich zu vereinfachen: Partnerinstitute können dazu in gewissem Umfang Druckwerke aus dem BSB-Bestand vor 1941<sup>25</sup> melden, die im jeweiligen Institut nicht vorhanden sind. Die BSB prüft die Titel auf Gemeinfreiheit, lässt im positiven Fall die Titel digitalisieren und stellt sie auf OstDok bereit. Das Institut kann daraufhin z. B. aus dem eigenen Katalog via Metadaten auf das Digitalisat auf OstDok verweisen und somit Hausangehörigen und Nutzern einen

---

<sup>21</sup> Das Bundesinstitut für ostwissenschaftliche und internationale Studien (BIOst) erforschte für Zwecke der Bundesregierung und zur Unterrichtung des Parlamentes die politischen, gesellschaftlichen und wirtschaftlichen Entwicklungen in der Sowjetunion, deren Nachfolgestaaten, den Staaten Ostmittel- und Südosteuropas, der Volksrepublik China und den kommunistisch regierten Staaten der Dritten Welt, sowie die Auswirkungen dieser Entwicklungen auf die internationalen Beziehungen. Es wurde Ende 2000 aufgelöst und ging in der SWP auf.

<sup>22</sup> [www.ssoar.info](http://www.ssoar.info)

<sup>23</sup> Dies wird möglichst durch Einbindung der Metadaten geschehen, die extern auf die Originaldaten zuführen.

<sup>24</sup> <http://www.priib.ru/Pages/Default.aspx>. Zur Kooperation siehe: <http://www.bsb-muenchen.de/Kooperation-mit-der-Russischen-Praesidentenbibliothek-Boris.3051.0.html>.

<sup>25</sup> 1941 ist das letztmögliche Erscheinungsdatum, wenn man berücksichtigt, dass Werke 70 Jahre nach dem Tod des Urhebers gemeinfrei werden. Die Jahresangabe ist dynamisch und wird während des Projekts jährlich hochgesetzt.

echten Mehrwert in der Literaturversorgung verschaffen. Die bisherigen Erfahrungen zeigen, dass ein ausgesprochen großes Interesse an dieser Möglichkeit besteht<sup>26</sup>.

Verlage zählen im Publikationsprozess – und in zunehmendem Maße auch im elektronischen – zu den wichtigsten Akteuren überhaupt. Das Verlagswesen selbst und erst recht der Umgang von Verlagen mit elektronischen Publikationen sind dabei als überaus heterogen zu charakterisieren: neben den zumeist größeren Verlagshäusern mit eigener Vertriebsplattform für e-books existieren nicht wenige Verlage ohne eigene Plattform für ihr e-book-Angebot bis hin zu solchen, die sich nach wie vor ausschließlich auf Printmedien stützen. Vornehmlich an die letztgenannten beiden Gruppen wendet sich dabei das Angebot, das eigene Verlagsprogramm über OstDok bekannt zu machen, indem die Publikationen – gegebenenfalls nach Ablauf einer vertretbaren Vermarktungsfrist (Moving Wall) – frei zugänglich auf OstDok bereitgestellt werden. Partnerverlage erhalten dazu auf OstDok eine eigene „Sicht“ zur Selbstdarstellung von Verlag und Programm, über die das Fachpublikum – und somit potentielle Käufer - wiederum zu einer externen Internetseite geführt werden kann<sup>27</sup>. Dabei können selbstverständlich verschiedenartig abgestufte Kooperationsmodelle Berücksichtigung finden<sup>28</sup>. Als Pilot dient im OstDok-Projekt zunächst der Sagner-Verlag, der alle Titel mit einer geringen Moving Wall im open access auf OstDok zur Verfügung stellen wird<sup>29</sup>.

Unter anderem in Kooperation mit der Hochschule für Medien in Stuttgart erfolgt aktuell eine Analyse des osteuroparelevanten Verlagswesens im deutschsprachigen Raum<sup>30</sup>: Welche Verlage sind für die Osteuropastudien besonders einschlägig und noch aktiv? Welche stehen dem open access-Gedanken nahe<sup>31</sup>?

---

<sup>26</sup> So bekundete bereits eine Vielzahl einschlägiger Einrichtungen Interesse an dieser Kooperationsform, darunter das Institut für Donauschwäbische Geschichte und Landeskunde in Tübingen (IDGL), <http://www.idglbw.de/>. Mit dem IDGL erfolgt aktuell ein entsprechendes Pilotprojekt.

<sup>27</sup> So gelangen Verlage in zunehmendem Maße zu der Einsicht, dass ein kostenloses Titelangebot im Internet nicht hinderlich, sondern mitunter auch förderlich für den Absatz der Printfassung eines Werkes sein kann.

<sup>28</sup> Vgl. als Grundlage hierzu etwa die im Rahmen von pedocs erarbeiteten Modelle; [http://www.pedocs.de/informationen\\_verlage.php?la=de](http://www.pedocs.de/informationen_verlage.php?la=de).

<sup>29</sup> Gleiches gilt auch für kommerzielle Reihen oder Zeitschriften, bei denen die Verwertungsrechte bei einem Herausgeber bzw. Herausbergergremium liegen. Eine derartige Übereinkunft besteht beispielsweise mit den Herausgebern des Wiener Slawistischen Almanachs, der künftig ebenfalls mit einer Moving Wall frei zugänglich erscheinen wird.

<sup>30</sup> Dies geschieht u.a. im Rahmen eines Projekts im Studiengang Bibliotheks- und Informationsmanagement der Hochschule für Medien in Stuttgart.

<sup>31</sup> Hinweise liefert dazu etwa die SHERPA/RoMEO-Liste (<http://www.sherpa.ac.uk/romeo/>), die Aufschluss darüber geben kann, welche Verlage welche Haltung bzgl. der Selbstarchivierung von Publikationen einnehmen.

Auf Grundlage der Ergebnisse gilt es, gezielt ausgewählte Verlage anzusprechen und für eine Zusammenarbeit im Bereich digitaler Bereitstellung zu gewinnen. Im Dialog mit Verlag und Urheber (in der Regel der Autor) sollen dabei explizit auch vergriffene Werke berücksichtigt werden, zu denen keine Neuauflage geplant ist<sup>32</sup>. Auf ähnliche Weise wird zu verfahren sein, wenn Verlagsrechte von „erloschenen“ Verlagen auf einen Rechtsnachfolger, in der Regel ebenfalls ein Verlag, übergegangen sind<sup>33</sup>.

Bei der technischen Entwicklung des OstDok-Repositoriums wird insbesondere mit der Max Planck Digital Library<sup>34</sup> und dem Fachinformationszentrum Karlsruhe<sup>35</sup> ein intensiver Austausch gepflegt. Deren modulare technische Systementwicklungen eSciDoc und Pubman bilden die Grundlage für das interne Datenverwaltungssystem von OstDok. Bei weiterer im Projekt verwendeter Open Source Software, die u.a. für den Suchindex eingesetzt wird, profitiert OstDok von der Entwickler-Gemeinschaft, gibt seinerseits aber auch Weiterentwicklungen an die Community zurück<sup>36</sup>.

Daneben bestehen in Entwicklungsfragen weitreichende Absprachen mit anderen Projekten, deren Plattformen an der Bayerischen Staatsbibliothek entstehen. Die Standardisierung reicht hier von der Präsentations- und Redaktionsoberfläche über den Datenfluss oder die Schnittstellen bis hin zum verwendeten Metadatenschema. Auf diese Weise wird sichergestellt, dass die technische Basis unmittelbar nachnutzbar ist, Datenaustausch ermöglicht wird und die Benutzerführung möglichst allgemeinen Standards entspricht - auch wenn OstDok nichtsdestotrotz eine Vielzahl spezieller Anforderungen umfasst<sup>37</sup>. Gerade die Usability stellt dabei einen wichtigen Erfolgsfaktor dar. Neben der kontinuierlichen Evaluation durch Nichtexperten wird OstDok auf diesem Gebiet künftig in hohem Maße von der Kooperation des Herder-

---

<sup>32</sup> In diesem Fall fällt das Verwertungsrecht zurück an den Urheber, der dann Ansprechpartner für die Einräumung von einfachen Nutzungsrechten ist. Daher sollen in einem weiteren Schritt die Autoren vergriffener Werke auch unmittelbar dazu aufgerufen werden, entsprechende Titel und Verlage zu melden, wenn Interesse an der elektronischen Bereitstellung besteht.

<sup>33</sup> Bei „erloschenen“ Verlagen steht zunächst die Ermittlung des Rechteinhabers im Mittelpunkt. Existiert dieser nicht, fallen auch hier die Verwertungsrechte an den Urheber zurück, der dann für eine Rechteeinräumung gewonnen werden soll.

<sup>34</sup> [www.mpd.l.mpg.de](http://www.mpd.l.mpg.de).

<sup>35</sup> [www.fiz-karlsruhe.de](http://www.fiz-karlsruhe.de).

<sup>36</sup> Beispielsweise wurde an der BSB ein Patch zum Ranking von Snippets beim Highlighten in der Trefferanzeige entwickelt und veröffentlicht; <https://issues.apache.org/jira/browse/LUCENE-3440>. In Zukunft sind insbesondere im Bereich der semantischen Suche und deren Optimierung für osteuropäische Sprachen Ergebnisse zu erwarten, die für einen breiten Anwenderkreis von Interesse sind.

<sup>37</sup> Hierzu zählt beispielsweise die Verarbeitbarkeit von mehreren Sprachen, Schriftsystemen und Transliterationen einschließlich der Sonderzeichen.

Instituts mit dem Zentrum für Medien und Interaktivität (ZMI) der Justus-Liebig-Universität Gießen profitieren. Im Rahmen dieser Zusammenarbeit wird eine Usability-Studie durchgeführt, deren Ergebnisse in die Weiterentwicklung von OstDok mit einfließen und dabei helfen sollen den Internet-Auftritt von OstDok weiter zu optimieren und den aktuellen Nutzerbedürfnissen und –verhalten anzupassen.

### **Vernetzung**

Gewährleisten die geschilderten Kooperationsformen im OstDok-Projekt in erster Linie den aktiven inhaltlichen und technischen Auf- und Ausbau der Plattform, so muss das Fachrepositorium in der Zukunft noch stärker mit anderen Distributionskanälen und best practise-Projekten vernetzt werden, um Expertenwissen auszutauschen, im Internet zusätzliche Einstiegsoptionen in die OstDok-Angebote zu schaffen oder allgemein den Bekanntheitsgrad von OstDok unter potentiellen Nutzern und Kooperationspartnern zu steigern.

Die Rezeption des Angebots hängt letztlich in hohem Maße davon ab, wie gut es gelingt, die Projekthinhalte bei Osteuropawissenschaftlern und Bibliothekaren zu kommunizieren. Die OstDok-Plattform wird daher kontinuierlich in Fachjournalen oder auf Tagungen, Kongressen oder Workshops vorgestellt. Daneben schaffen die Träger aber auch im Rahmen eigener Veranstaltungen Möglichkeiten der Vernetzung mit Einzelwissenschaftlern, Forschungs- und bibliothekarischen Einrichtungen. So richtete sich die 2010 vom HI in Zusammenarbeit mit der BSB organisierte Summer School „Neue Medien in den Geschichts- und Osteuropawissenschaften“ in erster Linie an den wissenschaftlichen Nachwuchs<sup>38</sup>. Ziel dieser Veranstaltung war es, die Teilnehmer gezielt einerseits für moderne Rechercheinstrumente zu sensibilisieren und andererseits für das Open Access-Publizieren zu interessieren. Eine stärker bibliothekarisch-informationswissenschaftliche Ausrichtung hatte dagegen der im Oktober 2011 veranstaltete Workshop "Digitale Volltexte und Hilfsmittel in den Osteuropastudien“, zu dem hauptsächlich Akteure und Mediatoren der Bereiche „Digitales Publizieren“ und „Bereitstellung“ an der Bayerischen Staatsbibliothek zusammentrafen.

---

<sup>38</sup> <http://www.herder-institut.de/startseite/struktur-und-aufgaben/sommerakademie/sommerakademie-2010.html>

Neben OstDok wurden hier zahlreiche weitere Projekte und Initiativen auch über den osteuropawissenschaftlichen Kontext hinaus präsentiert<sup>39</sup>. Künftig wird sich der Fokus der OstDok-Vermittlung zudem gezielt auf das Umfeld universitärer Lehre richten. Im Rahmen von Road-Shows soll das Projekt dazu in Kooperation mit den jeweiligen Universitätsbibliotheken vor Ort Lehrstühlen und Studierenden präsentiert werden.

Vor allem organisatorische Fragen zum eigentlichen Betrieb des Repositoriums – etwa zum Umgang mit Versionierungen von Dokumenten, zu urheberrechtlichen Problemfällen oder zur inhaltlichen Qualitätssicherung von genuin elektronischen Publikationen – lassen sich sinnvollerweise im Umfeld von Fachrepositorien diskutieren, die sich wegen eines früheren Projektbeginns bereits in einem weiteren Entwicklungsstadium befinden. Hierzu zählen im deutschsprachigen Raum insbesondere Plattformen wie econstor<sup>40</sup> (Wirtschaftswissenschaften), pedocs<sup>41</sup> (Erziehungswissenschaften), ssoar (Sozialwissenschaften) oder psydoc<sup>42</sup> (Psychologie). Eine geeignete Plattform für die Fachdiskussion rund um den Repositorienbetrieb stellt zudem die DINI<sup>43</sup>-Mailingliste Repositorymanagement dar<sup>44</sup>.

Mit dem Ziel, die Recherchierbarkeit von OstDok-Materialien im unübersichtlichen Internet zu verbessern, müssen für OstDok zusätzliche Distributionskanäle erschlossen werden, über die Nutzer zu den Inhalten der Plattform gelangen können. Dazu wurden und werden gewissermaßen auf allen Ebenen der vielschichtigen Informationsstruktur für Volltexte die OstDok-Metadaten in Rechercheinstrumente eingebunden. So sind sämtliche in OstDok erfassten Titel beispielsweise parallel im Katalog der Bayerischen Staatsbibliothek<sup>45</sup> recherchierbar und dadurch wiederum in sämtlichen Rechercheinstrumenten, die diesen Datenbestand nachnutzen, angefangen vom Verbundkatalog „B3Kat“<sup>46</sup> über den Karlsruher Virtuellen Katalog

---

<sup>39</sup> Die Veranstaltung wurde in Verbindung mit der Kommission für Fachreferatsarbeit des Vereins Deutscher Bibliothekare (VDB) durchgeführt und diente zugleich als Fortbildungsveranstaltung für FachreferentInnen.

<sup>40</sup> <http://www.econstor.eu/>.

<sup>41</sup> <http://www.pedocs.de/>.

<sup>42</sup> <http://psydok.sulb.uni-saarland.de/>.

<sup>43</sup> Deutsche Initiative für Netzwerkinformation, [www.dini.de](http://www.dini.de).

<sup>44</sup> <http://sympa.cms.hu-berlin.de/sympa/info/repositorymanagement>.

<sup>45</sup> <https://opacplus.bsb-muenchen.de/>.

<sup>46</sup> Hierbei handelt es sich um den gemeinsamen Verbundkatalog von Bibliotheksverbund Bayern und Kooperativem Bibliotheksverbund Berlin-Brandenburg mit über 23 Millionen Nachweisen.



(KVK)<sup>47</sup> bis hin zur weltweit größten bibliographischen Datenbank, dem WorldCat<sup>48</sup>. Insbesondere durch die Einbindung in den letztgenannten Universalkatalog wird eine enorme geographische Erweiterung des „Einzugsgebiets“ der Nutzer weit über die europäischen Grenzen hinaus erzielt und entsprechend der Bekanntheitsgrad der Plattform gesteigert.

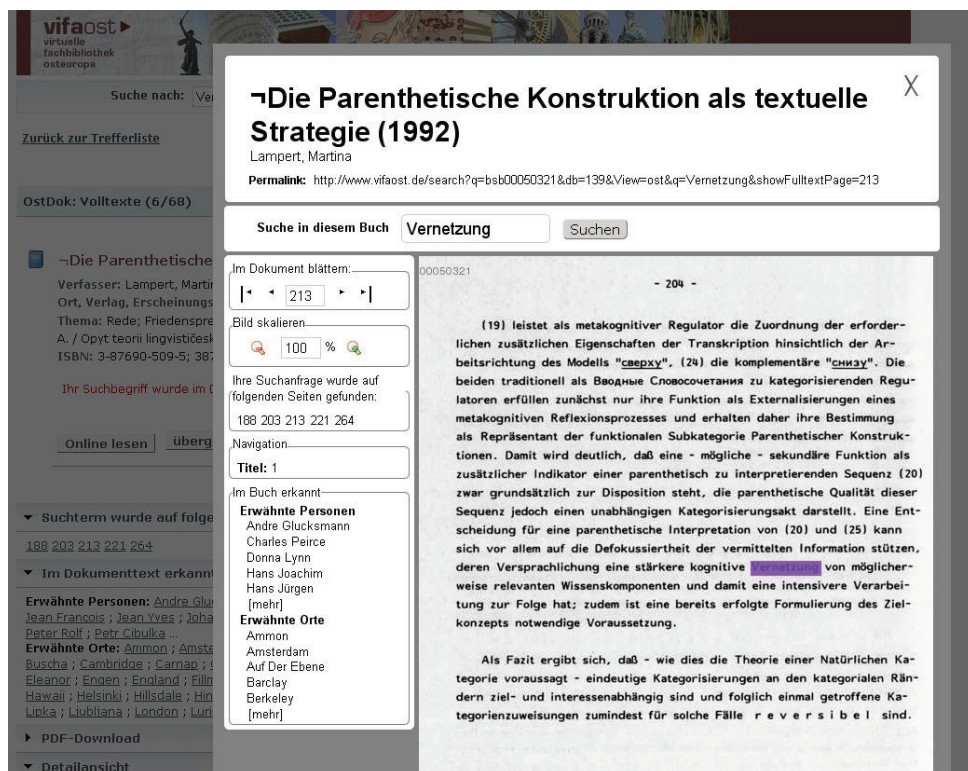


Abbildung 2: Volltexttrefferansicht in der ViFaOst-Metasuche, im Hintergrund. Einzeltrefferansicht

Ausgewiesenen Fachangeboten wie den Virtuellen Fachbibliotheken als „One-Stop-Shop“ für die jeweilige Disziplin kommt dort, wo sich Fachgebiete überschneiden, im Ausbau der Vernetzungsstruktur eine besondere Bedeutung zu. Besonders eng verzahnt ist die OstDok-Plattform bereits mit der Virtuellen Fachbibliothek Osteuropa

<sup>47</sup> Der KVK ist ein Metakatalog mit knapp 70 eingebundenen Katalogen von Bibliotheken und dem Buchhandel weltweit, <http://www.ubka.uni-karlsruhe.de/kvk.html>.

<sup>48</sup> <http://www.worldcat.org/>.

(ViFaOst), dem zentralen Portal für Fachinformationen aus dem Bereich der Osteuropastudien mit zahlreichen fachspezifischen Bibliothekskatalogen, E-Zeitschriften, Datenbanken, Volltextsammlungen, einem Veranstaltungskalender uvm. Interessierte Nutzer werden einerseits über einen direkten Zugang im Volltext-Modul der ViFaOst („Texte und Materialien“) zur OstDok-Plattform geleitet; andererseits lassen sich sämtliche in OstDok eingebundenen Publikationen über die ViFa-Metasuche recherchieren, freilich mit weniger komfortablen Suchmöglichkeiten als im Repositorium selbst (z.B. Browsingoption)<sup>49</sup>. Ähnliche Vernetzungen mit jeweils eigenen Schwerpunktsetzungen bieten sich dabei auch bei weiteren Virtuellen Fachbibliotheken wie etwa dem Slavistik-Portal<sup>50</sup> oder der vifanord<sup>51</sup> an.

Grundsätzlich stehen die Daten anderen Informationsdienstleistern zur freien Nachnutzung zur Verfügung, da der B3Kat seine Katalogdaten seit Dezember 2011 auch als Linked Open Data frei zugänglich anbietet<sup>52</sup>. Daneben gelangen OstDok-Daten aber gezielt auch auf anderen Wegen an bedeutsame Rechercheinstrumente. So liefert das für die Bereitstellung der OstDok-Publikationen verantwortliche Münchner Digitalisierungszentrums (MDZ) der Bayerischen Staatsbibliothek die Daten gezielt auch an ausgesuchte diversifizierende Angebote wie die wissenschaftliche Suchmaschine BASE (Bielefeld Academic Search Engine)<sup>53</sup> oder die europäische Virtuelle Bibliothek Europeana<sup>54</sup> aus.

Bilden die OstDok-Daten bei solchen Auslieferungen bislang noch jeweils einen Teilbestand des BSB-Gesamtbestandes, so wird OstDok nach der zeitnah geplanten

---

<sup>49</sup> <http://www.vifaost.de/>; <https://www.vifaost.de/metaopac/start.do?View=ost&SearchType=2>. In naher Zukunft wird die OstDok-Suche zudem unmittelbar in das Volltext-Modul integriert werden. Nach Möglichkeit sollen auch die in der ViFaOst vorhandenen Publikationen nachbearbeitet werden (Volltexterkennung und –korrektur) und in OstDok integriert werden.

<sup>50</sup> <http://www.slavistik-portal.de/>. Das Slavistik-Portal wird einen Fachausschnitt mit einschlägigen Volltexten in die eigene Volltextsuche integrieren.

<sup>51</sup> <http://www.vifanord.de/>. Die vifanord - Virtuelle Fachbibliothek Nordeuropa und Ostseeraum- wird OstDok als Volltextrepositorium für Publikationen zu den Baltischen Ländern nutzen. So werden einerseits Autoren mit ihren Publikationen an OstDok „vermittelt“, andererseits wird ein entsprechender Fachausschnitt in der vifanord bereitgestellt werden.

<sup>52</sup> Vgl. <http://www.bsb-muenchen.de/Archiv-Einzeldarstellung.395+M5c1fa0acc84.0.html>.

<sup>53</sup> <http://www.base-search.net>. BASE ist eine deutschsprachige akademische Suchmaschine mit einem Schwerpunkt auf open-access-Publikationen, die auf Dokumentenservern liegen. OstDok ist zudem in den beiden Repository-Registries OpenDOAR (Directory of Open Access Repositories, <http://www.opendoar.org/>) und ROAR (Registry of Open Access Repositories, <http://roar.eprints.org/>) registriert, die Open Access-Repositorien bzw. deren Inhalte nachweisen.

<sup>54</sup> <http://www.europeana.eu/portal/>.

Implementierung einer OAI-PMH-Schnittstelle<sup>55</sup> Daten auch als OstDok-Kollektion oder Teilmenge derselben direkt ausliefern können. Allgemeine Suchmaschinen wie google indexieren die OstDok-Metadaten dagegen schon jetzt unmittelbar selbst und sorgen somit für eine Präsenz in den weltweit meistgenutzten Rechercheinstrumenten.

Andere Repositorien rücken in den Vernetzungsfokus im Zusammenhang mit dem Komplex „Datentausch“. Vielfach bieten andere fachliche Repositorien sowie die Vielzahl institutioneller Repositorien (etwa von Hochschulen und Forschungseinrichtungen) Dokumente an, die auch für OstDok als einschlägig gelten. Wichtiger Bestandteil der Vernetzungsbestrebungen von OstDok wird daher in einer späteren Ausbaustufe auch der beiderseitige Austausch von Metadaten via Harvesting sein. Dabei müssen zunächst die technischen und inhaltlichen Mindeststandards der Publikationen und Bereitstellungsplattformen geprüft und langfristig sichergestellt werden. Die Einbindung fremder Inhalte kann je nach technischer Möglichkeit entweder nur die Nutzung der Metadaten oder auch die Indexierung der Volltexte umfassen. Von der Trefferanzeige in OstDok gelangt man dann durch Verlinkung zur Publikation auf der ursprünglichen Anbieter-Plattform. Die Rezeption dieser Materialien wird dadurch insgesamt gesteigert, was nicht nur im Interesse des ursprünglichen Bereitstellers, sondern auch des Autors liegt. Es ist daher damit zu rechnen, dass die Möglichkeiten des Harvestings auf breites Interesse stoßen wird. Entsprechende Absprachen mit einigen Fachanbietern liegen bereits vor. So sollen u.a. die Länderanalysen der Forschungsstelle Osteuropa<sup>56</sup> in Bremen ebenso eingebunden werden wie die osteuroparelevanten Angebote der Publikationsplattform „perspectivia.net“<sup>57</sup> oder Periodika der Plattform DIFMOE (Digitales Forum Mittel- und Osteuropa)<sup>58</sup>. In größtmöglichem Umfang sollen zudem Metadaten zu einschlägigen Digitalisaten auf universitären Repositorien eingeworben werden.

---

<sup>55</sup> Das Open Archives Initiative Protocol for Metadata Harvesting, kurz OAI-PMH, dient der einfachen Weitergabe von Metadaten zwischen Dokumentenservern und Service Providern. Anders als bei der verteilten Suche über die Z39.50-Schnittstelle, wo bei einer Anfrage die Datenbestände der einzelnen Anbieter einzeln angefragt und abgefragt wurden, werden bei OAI-PMH zunächst alle Metadaten eingesammelt, zentral indexiert und für die Suche bereitgestellt.

<sup>56</sup> <http://www.forschungsstelle.uni-bremen.de>.

<sup>57</sup> <http://www.perspectivia.net/>. Von Relevanz sind hier die Publikationsreihen der Deutschen Historischen Institute in Warschau und Moskau.

<sup>58</sup> <http://www.difmoe.eu/>.

## **Das Vernetzen von Menschen, Daten und Systemen – Die Forschungsbibliothek des Herder-Instituts in Marburg**

Jürgen Warmbrunn

Forschungsbibliothek Herder-Institut, Marburg

### **Zusammenfassung**

Dieser Beitrag möchte am Beispiel der seit dem Jahre 2000 als *Forschungsbibliothek* zur Geschichte, Kultur und Landeskunde Ostmitteleuropas<sup>1</sup> firmierenden größten organisatorischen Einheit des *Herder-Instituts für historische Ostmitteleuropaforschung – Institut der Leibniz-Gemeinschaft* in Marburg/Lahn anschaulich aufzeigen, dass das Vernetzen von Menschen, Daten und Systemen für diese Einrichtung seit jeher große Bedeutung hat. Insofern ist das, was gegenwärtig in Form einer vernetzten, partnerschaftlichen und kooperativen Zusammenarbeit mit vielen unterschiedlichen Akteuren den Kern ihrer Aktivitäten im bibliothekarischen sowie im Informationsbereich insgesamt bildet, im Grunde nichts anderes als eine konsequente Weiterentwicklung dessen, was die Arbeit der Bibliothek und ihrer Mitarbeiterinnen und Mitarbeiter in den letzten sechs Jahrzehnten bereits wesentlich bestimmt hat. Außerdem soll – allerdings mit aller bei der Darstellung zukünftiger Entwicklungsmöglichkeiten gebotenen Vorsicht – aufgezeigt werden, wie diese verschiedenen Formen der Vernetzung sich in den kommenden Jahren perspektivisch weiterentwickeln sollen. Konkret geht es dabei um eine Beschreibung möglicher Wege, die sie als größte Spezialbibliothek zu diesem Sammelgebiet im deutschsprachigen Raum mit einem bisher deutlich ausgeprägten Schwerpunkt im konventionellen Medienbestand beschreiten kann, um zusätzlich zu ihren bisherigen Dienstleistungen in einer über den streng bibliothekarischen Bereich hinausgreifenden Kooperation ein überzeugendes und an den Bedürfnissen ihrer Nutzerinnen und Nutzer orientiertes Produktfolio von elektronischen bzw. virtuellen Angeboten aufzubauen.

### **Abstract**

This contribution shows that the networking of people, data and systems has always played an important role for the Library of the Herder-Institut in Marburg. Its present tight and target-oriented cooperation with a huge number of different partners is therefore nothing but a consistent continuation of a cooperative approach that has characterized the work of the library and its staff over the last six decades. In addition it will be shown how the different forms of networking already practiced are likely to

---

<sup>1</sup> Ostmitteleuropa wird hier verstanden als das Gebiet der heutigen Staaten Estland, Lettland, Litauen, Polen, Slowakei und Tschechien sowie der russischen Exklave Kaliningrad.

develop further. This will include an analysis of how the library can in future in the interest of its users develop a product folio of electronic or virtual resources that will supplement its present excellent print holdings, which already make it the largest special library on the history and culture of East Central Europe (Poland, the Czech Republic, Slovakia, Estonia, Latvia, Lithuania and the Russian exclave Kaliningrad) in the German-speaking countries.

### **Wie alles begann**

Als das Herder-Institut im Jahre 1950 in Marburg gegründet wurde, geschah dies mit dem Ziel eines bewussten und gezielten *Networkings*, auch wenn diesen Begriff vermutlich keiner der Gründungsväter der neuen Einrichtung jemals verwandt hat. Vernetzt werden sollten zunächst einmal Personen, nämlich vor allem die Mitglieder ihrer Trägereinrichtung, des Johann-Gottfried-Herder-Forschungsrats, eines akademieähnlichen Zusammenschlusses von Wissenschaftlern (von Wissenschaftlerinnen war noch weniger die Rede) aus den ehemaligen deutschen Ostgebieten und deutschen Siedlungsgebieten im östlichen Europa, die in Folge des Zweiten Weltkriegs häufig ihre wissenschaftliche, aber auch ihre ganz konkrete geographische Heimat verloren hatten. Das neu gegründete Institut sollte die institutionelle Basis für diesen Zusammenschluss bilden, sie bei ihrer Arbeit unterstützen und für eine Verbreitung der Ergebnisse ihrer wissenschaftlichen Arbeit sorgen.

Auch wenn die vergleichsweise günstige Bibliothekssituation in Marburg – die dortige Universitätsbibliothek hatte keine nennenswerten Kriegsverluste erlitten, außerdem wurden seit 1946 in Marburg die in die amerikanische Besatzungszone verlagerten Bestände der Berliner Staatsbibliothek zunächst als „Hessische Bibliothek“, später als „Westdeutsche Bibliothek“ und schließlich als „Staatsbibliothek Preußischer Kulturbesitz“ der Öffentlichkeit zugänglich gemacht – vermutlich einer der wesentlichen Gründe war, das Herder-Institut in der Universitätsstadt an der Lahn anzusiedeln, war ein solches Institut aus damaliger Sicht ohne eine eigene Bibliothek überhaupt nicht vorstellbar. Entsprechend wurde unmittelbar mit dem Aufbau einer eigenen geistes- und kulturwissenschaftlich ausgerichteten Bibliothek zur Geschichte und Gegenwart Ostmitteleuropas – damals freilich noch mit einem ganz dezidierten Fokus auf dem deutschen Anteil an dessen bisheriger Entwicklung – begonnen.

Vernetzung war auch für die neu gegründete Bibliothek ein entscheidendes Kriterium: So begann man etwa sehr rasch mit der klassischen Formen der Vernetzung im bibliothekarischen Erwerbungsgeschäft, dem Tausch: Nur ein Jahr nach ihrer Gründung nahm die Bibliothek bereits den Tausch mit der Library of Congress in Washington, D.C., auf und selbst über die Grenzen der beiden großen Machtblöcke in Europa hinweg bestanden beispielsweise schon seit 1956 Kontakte mit der Universitätsbibliothek Wrocław/Breslau. Schon früh wurde aber auch ein weiteres Vorhaben der Vernetzung in Angriff genommen: Die Zusammenführung der Nachweise von Bibliotheksbeständen, die für die Beschäftigung mit Ostmitteleuropa im Herder-Institut, aber nicht nur dort als unbedingt notwendig empfunden wurden. Am Herder-Institut entstand auf diese Weise seit 1950 der „Gesamtkatalog Ostmitteleuropa“, ein konventioneller Kartenkatalog, der Bestände in rund 33 Bibliotheken zusammenfasste und auf der Lieferung von duplizierten Katalogkarten dieser Bibliotheken beruhte. Diese wurden in Marburg in einem separaten Katalog zusammengeführt.<sup>2</sup> Als fachlich orientierte Entsprechung zu den regionalen Zentralkatalogen sollte dieser Katalog im vorelektronischen Zeitalter durch den verlässlichen Nachweis einschlägiger Bestände den bibliothekarischen Leihverkehr erleichtern und dies interessanterweise nicht nur auf nationaler Basis (wie bei den regionalen Zentralkatalogen), sondern durchaus unter bewusster Einbeziehung und Mitwirkung von Bibliotheken mit Spezialbeständen zu diesem Bereich im Ausland.<sup>3</sup>

Diese kollaborative und gewissermaßen am bibliothekarischen bzw. wissenschaftlichen Gemeinwohl orientierte Arbeit wurde bis in die 1990er Jahre fortgesetzt, lange nachdem die Bibliothek des Herder-Instituts viel weitergehende Schritte in Richtung Vernetzung unternommen hatte: Seit 1977 beteiligte sie sich an der Erstellung eines Hessischen Zeitschriftenverzeichnisses, 1984 erfolgte die Übernahme des damals neuen Regelwerks RAK-WB und 1989 der Eintritt in den EDV-gestützt arbeitenden HeBIS-Bibliotheksverbund, der dann im Herbst 1995 zum PICA-Bibliothekssystem migrierte. Ebenfalls seit 1995 katalogisierte die Bibliothek

---

<sup>2</sup> Ein weiterer derartiger Zentralkatalog entstand in der Nachkriegszeit in Form des „Osteuropa-Zentralkatalogs“ der Osteuropa-Abteilung der Staatsbibliothek Preußischer Kulturbesitz. Anders als im Herder-Institut wurde dieser nicht durchgehend als eigenständiger zusätzlicher Bestandskatalog geführt.

<sup>3</sup> Teilnehmer am Gesamtkatalog Ostmitteleuropa waren u.a. die Universitätsbibliothek Wien oder - mit ihrem Altbestand - die Königliche Bibliothek Kopenhagen. Dies galt auch für die Universitätsbibliothek Helsinki mit ihren hervorragenden, auf der Pflichtexemplar-Ablieferung gegründeten historischen Osteuropa-Beständen aus der Zeit der Zugehörigkeit Finnlands zum Zarenreich. Eine moderne Fortsetzung fand der Gesamtkatalog Ostmitteleuropa im von der Martin-Opitz-Bibliothek Herne betreuten Verbundkatalog Östliches Europa (VOE), an dem sich das Herder-Institut jedoch nicht beteiligt. Vgl. dazu <http://www.martin-opitz-bibliothek.de/de/voe-info/>.

ihre Zeitschriften in der Berliner Zeitschriften-Datenbank und zählte somit zu den ersten Spezialbibliotheken, die hier ihre Zeitschriftennachweise führte.

Dieser stark kooperative Zugang hatte seinen Grund vermutlich auch in der Tatsache, dass die Bibliothek des Herder-Instituts nach ihrer frühen Erwähnung in entsprechenden Empfehlungen des Wissenschaftsrats aus dem Jahre 1964 (Empfehlungen 1964: 224 ff.) schon bald durch die Deutsche Forschungsgemeinschaft in den Kreis der durch sie geförderten Spezialbibliotheken von überregionaler Bedeutung aufgenommen wurde.<sup>4</sup> Damit einher ging eine weitere deutliche Öffnung der Bibliothek für die externe Nutzung und zwar nicht nur vor Ort, sondern auch im Rahmen des auswärtigen Leihverkehrs, an den sie 1971 angeschlossen wurde.

Die Bibliothek des Herder-Instituts verstand und versteht ihre Mitarbeit im Bibliotheksverbund und beim Aufbau der früher getrennten, heute in Form der GND zusammengefassten Normdateien in letzter Konsequenz als kollaborative Unterfangen, die frühere Aktivitäten wie etwa die Führung des „Gesamtkatalogs Ostmitteleuropa“ unter neuen organisatorischen und technischen Rahmenbedingungen fortsetzen. Dieser Aspekt der Vernetzung im Sinne des „bibliothekarischen Gesamtinteresses“ lässt es für ein vergleichsweise kleines, aber stark spezialisiertes wissenschaftliches und Infrastrukturinstitut vertretbar erscheinen, neben dem Nutzen der Verbundkatalogisierung auch die Lasten einer qualitativ hochwertigen Verzeichnung beispielsweise von Zeitschriften oder einer möglichst anspruchsvollen Ermittlung von Normdaten zu Personen auf sich zu nehmen<sup>5</sup>. Aus dieser Motivation heraus übernahm die Bibliothek des Herder-Institut vor einigen Jahren auch die Aufgabe, im Rahmen der Einführung des originalschriftlichen Katalogisierens im HeBIS-Verbund als Pilotprojekt für die Katalogisierung mit kyrillischen Schriftzeichen zu fungieren.

---

<sup>4</sup> Diese Förderung wurde der Bibliothek bis zur Einstellung der entsprechenden Förderlinie zum Jahresende 2010 ununterbrochen zuteil.

<sup>5</sup> Als Direktteilnehmerin an der ZDB katalogisiert die Bibliothek grundsätzlich alle Zeitschriftentitel selbst, als Mitarbeiterin an der bisherigen Personennamendatei mit höchster Redaktionsstufe nimmt sie insbesondere bei Personen aus Ostmitteleuropa grundsätzlich möglichst weitgehende Individualisierungen und Ergänzungen der die jeweilige Person identifizierenden Detailinformationen vor.



Ein weiteres Indiz für die stark internationale Vernetztheit der Bibliothek des Herder-Instituts bildet im Übrigen der Nachweis ihrer Bestände im WorldCat von OCLC, auf den an anderer Stelle im Zusammenhang mit der angestrebten Georeferenzierung aller Bestände des Herder-Instituts noch näher einzugehen sein wird.

### **Veränderungen daheim, in Ostmitteleuropa und in der Welt**

Die Tatsache, dass das Herder-Institut zu Beginn des Jahres 1994 rechtlich verselbstständigt worden war und im Rahmen der gemeinsamen Forschungsförderung des Bundes und der Länder in seiner Satzung den Auftrag erhalten hatte, als „wissenschaftliche Serviceeinrichtung [...] die Erforschung der Länder und Völker des östlichen Mitteleuropa in europäischen Bezügen“ zu unterstützen, hatte für die Bibliothek insofern nur geringe Auswirkungen, als gerade dort der Dienstleistungscharakter ohnehin seit jeher das Tagesgeschäft bestimmt hatte. Eher schon war die politische Wende in Ostmitteleuropa der Jahre 1989/90 für die bibliothekarische Arbeit von Bedeutung, wurden doch Aufenthalte von Gästen aus dem Arbeitsgebiet des Instituts dadurch ebenso einfach wie Gegenbesuche, wobei letzteres ganz besonders für die ehemaligen baltischen Sowjetrepubliken zutraf, wo etwa im Falle Estlands die altherwürdige und bedeutende Universitätsstadt Tartu/Dorpat aufgrund der Nähe einer sowjetischen Luftwaffenbasis lange eine für Ausländer völlig unzugängliche Stadt gewesen war. Zeitweise Unterbrechungen im Tauschverkehr konnten ebenso bewältigt werden wie die Wirrungen, die sich bei der Medienbeschaffung zumindest übergangsweise aus der Transformation der bisherigen staatlichen Verlage und Buchhandlungen bzw. dem Entstehen neuer Verlage und Anbieter ergaben.

Nur kurze Zeit nach den Revolutionen in Ost-, Ostmittel- und Südosteuropa erfolgte in der ersten Hälfte der 1990er Jahre eine weitere unblutige Revolution, diesmal allerdings im Informationsbereich, die vielleicht durch den Begriff des World Wide Web am besten charakterisiert wird. Die Veränderungen im Kommunikations- und Informationsverhalten, die zu diesem Zeitpunkt ihren Anfang nahmen, erreichten auch das Herder-Institut und seine Bibliothek sehr schnell und führten zu einer partiellen Neuausrichtung ihrer Arbeit, die bis heute anhält.



### **Erste virtuelle Angebote**

Da – auch unter dem Aspekt der jeweiligen Förderung durch die Deutsche Forschungsgemeinschaft – bereits seit langer Zeit enge Kontakte zwischen der Osteuropa-Abteilung der Bayerischen Staatsbibliothek in München und dem Herder-Institut bestanden hatten, ist es nicht überraschend, dass das Herder-Institut zu einem der ersten Partner wurde, als es darum ging, im Rahmen der Förderlinie „Verteilte Digitale Forschungsbibliothek“ der DFG auf den sich immer verstärkenden medialen Wandel zu reagieren und im Sinne eines modernisierten Sondersammelgebietsauftrags neben gedruckten Bibliotheksbeständen zunehmend auch virtuelle fachspezifische Angebote für die Öffentlichkeit zur Verfügung zu stellen. Das Ergebnis dieser Zusammenarbeit ist die Virtuelle Fachbibliothek Osteuropa (ViFaOst), die seit 2001 online ist und auch nach Auslaufen der DFG-Finanzierung durch die ursprünglichen sowie zwei neue Partner aus Eigenmitteln vollumfänglich aufrecht erhalten wird. Darüber hinaus ergab sich für das Herder-Institut auch eine punktuelle Zusammenarbeit mit zwei weiteren virtuellen Fachbibliotheken: Insbesondere mit Blick auf das Baltikum mit der Virtuellen Fachbibliothek Nordeuropa und Ostseeraum (vifanord) sowie dem Slavistik-Portal als einem eher sprach- und literaturwissenschaftlich ausgerichteten Angebot der Berliner Staatsbibliothek. Im Rahmen der ViFaOst bietet das Herder-Institut u.a. seine bibliographische Datenbank an und betreut gemeinsam mit der Bayerischen Staatsbibliothek den Katalog der Internetressourcen OstNet zu Ost-, Ostmittel- und Südosteuropa (hier mit Zuständigkeit für Ostmitteleuropa). Auch die Erfassung und Erschließung für die Ostmitteleuropaforschung relevanter Internetquellen erfolgt dabei seit einigen Jahren im Rahmen eines kooperativen Netzwerks: Mehr als 35 Einrichtungen mit über 120 Redakteuren beteiligen sich an dem seit 2001 entwickelten Verbund „Academic Link Share“<sup>6</sup>.

Eng verknüpft mit der Virtuellen Fachbibliothek Osteuropa wird in Zukunft ein weiteres Angebot sein, das die Bayerische Staatsbibliothek in Zusammenarbeit mit dem Herder-Institut und zusätzlichen Partnern 2009 wiederum im Rahmen eines DFG-Projekts in Angriff nehmen konnte. Es handelt sich dabei um "Osteuropa-Dokumente online" (OstDok), eine zentrale Plattform für elektronische Volltexte der Osteuropaforschung. In Form einer engen Zusammenarbeit mit der BSB als DFG-

---

<sup>6</sup> Mehr dazu unter <http://www.academic-linkshare.de/>

Sondersammelgebietsbibliothek und wissenschaftlichen Einrichtungen der Osteuropaforschung soll ein Fachrepositorium für Osteuropastudien entstehen, das wissenschaftlich relevante Volltexte nach dem Prinzip des Open Access im Netz nutzbar und durchsuchbar zur Verfügung stellt. Hierbei geht es in der aktuellen Anfangsphase des Projekts primär um retrodigitalisierte Texte – neben Publikationen der beteiligten Institute auch um zusätzliche Veröffentlichungen von Projektpartnern, „eingekaufte Monographien“<sup>7</sup> und Retrodigitalisate, die im Rahmen der Public-Private-Partnership zwischen der Bayerischen Staatsbibliothek und Google entstehen. In einer zweiten Phase soll OstDok darüber hinaus aber auch die Möglichkeit zu einem eigenständigen Upload von Texten durch ihre Verfasser bieten. Diese Publikationen sollen dann nach entsprechender Prüfung ebenfalls online zur Verfügung gestellt werden. Dem Ziel, auf diese Weise das wissenschaftliche Publizieren im Sinne des Open Access zu fördern und gleichzeitig junge Wissenschaftlerinnen und Wissenschaftler zu animieren, auch hochwertige Publikationen wie z.B. sehr gut bewertete Qualifikationsarbeiten in elektronischer Form zu veröffentlichen, entspricht OstDok durch die Einrichtung von eigenen Reihen mit in der Osteuropaforschung renommierten Herausgebern und einem anspruchsvollen Redaktionsverfahren, wie es heute auch bei Printpublikationen eher die Ausnahme als die Regel darstellt.

### **Bibliographische Arbeiten**

Das Herder-Institut begann bald nach seiner Gründung im Jahre 1950 damit, bibliographische Arbeiten zur Geschichte Ostmitteleuropas zu unterstützen bzw. auch selber durchzuführen. Dies geschah im Einklang mit der ursprünglichen Schwerpunktsetzung der Institutsarbeit auf der Geschichte und Kultur der Deutschen im östlichen Mitteleuropa. So entstanden – teilweise in Zusammenarbeit mit den ostdeutschen Historischen Kommissionen – historische Regionalbibliographien zu den ehemaligen deutschen Ostgebieten, die das Angebot der in der Bundesrepublik und der DDR erscheinenden jeweiligen regionalen Bibliographien ergänzten. In der 2. Hälfte der 1990er Jahren wurde die bibliographische Arbeit im Herder-Institut konzentriert und die Erfassung der Daten EDV-gestützt fortgeführt.

---

<sup>7</sup> Beispielsweise im Rahmen des Digi20-Projekts. Vgl. <http://digi20.digitale-sammlungen.de/>

Um das Jahr 1997 entstand mit der Literaturdokumentation eine eigene bibliographische Abteilung, die in Kooperation mit Partnern insbesondere in Ostmitteleuropa eine Allegro-Datenbank zur Geschichte Ostmitteleuropas betrieb und daraus insgesamt sieben gedruckte Jahresbibliographien generierte. Obwohl die Literaturdokumentation u.a. die Titelaufnahmen der Bibliothek des Herder-Instituts nachnutzen konnte und eine Arbeitsteilung mit ihren ausländischen Partnern betrieb, erwies sich die bibliographische Arbeit und hier insbesondere die Herausgabe gedruckter Bibliographien als so zeit- und ressourcenaufwändig, dass letztere mit immer größerer zeitlicher Entfernung zum Berichtszeitraum erschienen<sup>8</sup> und schließlich die Entscheidung getroffen wurde, weitgehend auf die Herausgabe gedruckter Jahresbibliographien zu verzichten.

Nicht zuletzt in Zusammenhang mit einer Empfehlung im Evaluierungsbericht des Jahres 2006 wurde im Herbst 2009 eine externe Expertengruppe eingesetzt, die eine enge Zusammenarbeit der Abteilungen Bibliothek und Literaturdokumentation sowie eine Reduzierung der Vielzahl der Datenbankanwendungen im Institut empfahl. Nach einer zusätzlichen Evaluierung der externen Kooperationsbeziehungen durch die Institutsleitung und in Absprache mit den Institutsmitgliedern wurde die Literaturdokumentation als Arbeitsbereich Bibliographieportal zum 1.7.2011 in die Forschungsbibliothek integriert und zugleich damit begonnen, die Migration der Literaturdatenbank in das Zentralsystem (CBS) des hessischen Bibliotheksverbunds HeBIS in die Wege zu leiten. Diese Migration soll nach Möglichkeit bis zum Jahresende 2012 abgeschlossen werden.

Da die Literaturdokumentation neben dem Bestand der Forschungsbibliothek – bei Zeitschriften und Sammelwerken bis auf Aufsatzebene – auch weitere osteuropabezogene Literatur verzeichnet, die nicht im eigenen Haus vorhanden ist, stellte sich im Vorfeld der Datenmigration und -zusammenführung im Rahmen des HeBIS-Verbundes sehr rasch die Frage nach einer möglichst nutzungsfreundlichen Präsentation. Hier wurde entschieden, statt der ursprünglich geplanten zwei getrennten OPACs – einer für den physischen Bestand der Forschungsbibliothek, einer für den gesamten bibliographischen Datenbestand – nur einen gemeinsamen OPAC anzubieten, bei dem jedoch auf allen Ebenen klar zwischen physisch

---

<sup>8</sup> Die einzige Ausnahme bildete hier die von Paul Kaegbein verantwortete und in Zusammenarbeit mit der Baltischen Historischen Kommission herausgegebene Baltische Bibliographie.

vorhandenen bzw. nur bibliographisch nachgewiesenen Titeln unterschieden werden kann. Eine weitere – im Hinblick auf die internationale Vernetzung des Herder-Instituts möglicherweise noch weitere – Frage ergab sich im Zusammenhang mit den bibliographischen Kooperationspartnern des Instituts, mit denen zum Teil seit vielen Jahren ein Datenaustausch betrieben wird. Für das Herder-Institut sehr erfreulich war hier die Aussage der HeBIS-Verbundzentrale, sowohl den Datenimport als auch den Datenexport grundsätzlich unabhängig vom Datenformat anbieten zu wollen, auch wenn die Zusammenarbeit insbesondere mit bibliothekarischen Partnern, die das MARC-Format oder andere gängige bibliothekarische Bibliotheken anbieten, naturgemäß wesentlich weniger zeit-, ressourcen- und kostenaufwändig sein wird. Hier wird das Herder-Institut somit vom Übergang des deutschen Bibliothekswesens zu einer stärker vernetzten internationalen Arbeit profitieren können.

Wichtige weitere Aspekte der Zusammenführung zweier bisher getrennt arbeitender Abteilungen werden in den kommenden Monaten geklärt werden müssen: In welchem Umfang können nachweisliche Vorzüge der bisherigen Allegro-basierten Literaturdokumentation (z.B. leichte Umsetzung von Mehrsprachigkeit) unter den neuen Rahmenbedingungen aufrecht erhalten werden, welche Auswirkungen hat die verbundgestützte Arbeit sowohl für den – bibliothekarischen wie bibliographischen – Workflow als auch für die Kooperationsbeziehungen und in welchem Maße bringen Verbunddienstleistungen (Fremddaten) bzw. eigene Angebote (Kataloganreicherung) eine Optimierung der bibliographischen Arbeit mit sich?

Die seit dem Jahre 2011 in der Bibliothek erfolgende Kataloganreicherung durch Scans der Inhaltsverzeichnisse aller nicht-deutschsprachigen Neuzugänge sowie konsekutiv auch des Altbestands ist ein weiteres Beispiel für eine nicht ausschließlich an den Hausbedürfnissen orientierte Dienstleistung der Bibliothek des Herder-Instituts. Im Einklang mit mehreren weiteren Bibliotheken des HeBIS-Verbundes nutzt das Herder-Institut intelligentCAPTURE der Firma AGI Information Consultants und ist mit seinen Beständen darüber hinaus in der Suchmaschine für wissenschaftliche Literatur *dandelion.com* verzeichnet. Ausgangspunkt für den Einstieg in die Kataloganreicherung war dabei eine über den bibliothekarischen Bereich hinausgehende Vernetzung des Herder-Instituts mit der Justus-Liebig-Universität Giessen und dort insbesondere dem Giessener Zentrum Östliches Europa (GiZO). Beim GiZO handelt es sich um ein regionalwissenschaftliches

Forschungs- und Lehrzentrum zum östlichen Europa, das im Juli 2006 gegründet wurde und die geisteswissenschaftliche Forschung und Lehre zum östlichen Europa in Hessen bündelt<sup>9</sup>. Dem vorausgegangen war eine Verlagerung der Fächer Osteuropäische Geschichte und Slavistik aus Marburg nach Giessen und damit verbunden eine – in der gemeinsamen Berufung des Direktors des Herder-Instituts mit der Universität Giessen besonders deutlich werdende – Neuorientierung des Herder-Instituts in Richtung Giessen, ohne dass dabei jedoch die jahrzehntelange Zusammenarbeit mit der Marburger Universität aufgegeben wurde. Nachdem die Giessener Universitätsbibliothek bereits begonnen hatte, ihre osteuropabezogenen Neuzugänge mit Kataloganreicherungsdaten zu versehen, zog die Bibliothek des Herder-Instituts nach und zwar ausdrücklich auch im Interesse der Giessener Studierenden und Mitarbeiter, denen bei der OPAC-Suche zusätzliche Entscheidungshilfen zur Relevanz der im Herder-Institut vorhandenen Titel gegeben werden sollten. Dies geschah wiederum vor einem ganz praktischen Hintergrund: Angehörige der Universität Giessen erhalten auf Wunsch Benutzungsausweise der Bibliothek des Herder-Instituts und können bei Bedarf im OPAC des Herder-Instituts Titel zur Ausleihe nach Giessen bestellen, wo sie diese kostenfrei an der Ausleihtheke der dortigen Universitätsbibliothek erhalten.<sup>10</sup> Dieser sogenannte „privilegierte Nutzerservice“ zur bevorzugten und schnellen Versorgung mit Literatur des Spitzenbedarfs weckte im Bereich der historischen Ostmitteleuropaforschung so viel Aufmerksamkeit, dass er mittlerweile auch für das Imre Kertész Kolleg Jena an der Friedrich-Schiller-Universität Jena eingeführt wurde und voraussichtlich ab Herbst 2012 auch mit dem Geisteswissenschaftlichen Zentrum Geschichte und Kultur Ostmitteleuropas (GWZO) an der Universität Leipzig vereinbart werden wird.

### **E pluribus unum oder Aus vielen weniger**

Bedingt durch die Vielfalt und Vielzahl der im Herder-Institut versammelten und zu erschließenden Materialien auf der einen Seite und den frühen Beginn der elektronischen Verzeichnung dieser Bestände in unterschiedlichen Datenbanken und Verzeichnissen auf der anderen Seite wurde bereits nach der Jahrtausendwende immer stärker die Frage diskutiert, wie man sowohl im Nutzerinteresse als auch zur

---

<sup>9</sup> Ebenso wurde die hessische auf Ostasien bezogene universitäre Forschung und Lehre in Frankfurt/Main und die zum Orient in Marburg gebündelt.

<sup>10</sup> Ein analoges Angebot besteht auch seitens der Universitätsbibliothek Giessen für die Mitarbeiter des Herder-Instituts.

Vermeidung überflüssiger Doppel- bzw. Mehrfacharbeiten zu einer Verschlankung der bestehenden Verzeichnisstruktur und idealerweise zur Integration von Datenbanken kommen könnte. Dies führte zu der bereits beschriebenen Empfehlung einer externen Expertengruppe, die bibliographische Datenbank in den HeBIS-Bibliotheksverbund zu integrieren und die Abteilungen Bibliothek und Literaturdokumentation nach Möglichkeit auch organisatorisch zusammenzufassen. Deutlich schwerer tat sich die Expertengruppe hingegen mit Empfehlungen zu einer Vereinheitlichung der in den Wissenschaftlichen Sammlungen und hier insbesondere dem Bildarchiv sowie der Dokumentensammlung, dem Archiv des Herder-Instituts, vorgehaltenen Verzeichnisse und Datenbanken.<sup>11</sup>

Angesichts der eher generellen Empfehlungen der externen Expertengruppe zu diesem Bereich wurde nach längeren Beratungen einer entsprechenden Arbeitsgruppe im Institut schließlich konkret entschieden, die Struktur der bereits vorhandenen Datenbanken weitestgehend beizubehalten, als Vorstufe für ein zukünftiges Fachinformationsportal zu Ostmitteleuropa bzw. eine einschlägige Forschungsumgebung des Herder-Instituts aber eine deutliche Verbesserung der Suchmöglichkeiten durch eine Erschließung mit einheitlichen Deskriptoren anzustreben. Als sinnvollsten ersten Schritt auf diesem Wege identifizierte man eine einheitliche Erschließung der in den verschiedenen Beständen relevanten Personen, auf die in einem zweiten späteren Schritt eine – gerade im ostmitteleuropäischen Kontext mit seinem häufigen Wechsel der Ortsnamenbezeichnungen ausgesprochen wünschenswerte – gemeinsame Georeferenzierung der Ortsbezeichnungen folgen sollte. Als probates Mittel bei der Erschließung von Personen wurde die damalige Personennamendatei (seit dem 19. April 2012 aufgegangen in der Gemeinsamen Normdatei) identifiziert, die sich mittlerweile bereits weit über ihren ursprünglichen bibliothekarischen Anwendungsbereich verbreitet hat und beispielsweise auch in Online-Portalen Anwendung findet. Entsprechend wurde Anfang 2012 mit dem Aufbau eines Zentralen Personenregisters begonnen, in dem – beginnend mit den Personendaten der Forschungsbibliothek – alle im Herder-Institut vorhandenen

---

<sup>11</sup> Die Bestände der Kartensammlung sind – soweit es sich um gedruckte Materialien handelt – bereits im Elektronischen Katalog der Bibliothek bzw. in der IKAR-Altkartendatenbank (Datenbank der Altkartendrucke vor 1850 beim GBV) verzeichnet. Vgl. dazu <http://ikar.staatsbibliothek-berlin.de>. Die Bestände der Pressesammlung des Herder-Instituts auf Titlebene sind ebenfalls im OPAC der Forschungsbibliothek bzw. in der ZDB nachgewiesen. Dies gilt im Übrigen auch für weite Teile der zum Teil sehr seltenen bzw. unikalen Bestände der Musiksammlung (*musica practica* und *musica theoretica*), die ebenso wie die Pressesammlung auch organisatorisch Teil der Forschungsbibliothek ist.

Daten zu Personen zusammengeführt, identifiziert, ggf. miteinander abgeglichen, bei Bedarf individualisiert und schließlich unter einer einheitlichen Suchoberfläche, die durch den Arbeitsbereich Informationstechnologie am Institut eigenständig entwickelt wird, zur Verfügung gestellt werden sollen.

In einem weiteren Schritt soll für alle Sammlungsbereiche die Georeferenzierung als Unterstützung bei der Suche genutzt werden. Beispiele für deren praktische Anwendung finden sich im unmittelbaren Arbeitsumfeld des Herder-Instituts etwa in der Virtuellen Fachbibliothek Nordeuropa und Ostseeraum (vifanord), der das Herder-Institut bereits als Kooperationspartner verbunden ist. Schon seit mehreren Jahren bietet die vifanord für die Baltischen Länder einen geographischen Sucheinstieg in Form einer Beta-Version.<sup>12</sup> Einen anderen Ansatz verfolgt das finnische Angebot „Kulttuurisampo – suomalaisen kulttuuri semantisessa Web 2.0:ssa“ [Kultursampo<sup>13</sup> – finnische Kultur im semantischen Web 2.0], ein Modell, an dem sich das Herder-Institut beispielsweise bei seiner Übersicht über die Präsentationsorte der Ausstellungen des Herder-Instituts seit dem Jahr 2000 orientiert hat.<sup>14</sup> Anders als bei der Kartensuche in „Kulttuurisampo“ werden die Detailinformationen zu den Ausstellungen des Herder-Instituts auf dem Eingangsbildschirm aber erst nach Anklicken der jeweiligen Orte präsentiert, was letztlich ein möglicherweise verstörendes Überangebot an Informationen verhindert.

In diesem Zusammenhang ebenfalls zu erwähnen ist das Projekt *mapfast* von OCLC, dem das Herder-Institut durch seine Mitgliedschaft im hessischen HeBIS-PICA-Verbund und im WorldCat verbunden ist. Diese bisher nur in Form eines Prototyps zugängliche Suche basiert auf einem Google Maps-Interface und führt bei Recherchen zur Geschichte Ostmitteleuropas in Deutschland bereits heute häufig zu Exemplaren der Bibliothek des Herder-Instituts als nächstgelegenen Treffern.<sup>15</sup>

Eine erhebliche synergetische Weiterentwicklung georeferenzierter Angebote des Herder-Instituts wird in den kommenden Jahren das im Rahmen der eHumanities-Ausschreibung des Bundesministeriums für Bildung und Forschung bewilligte Projekt

---

<sup>12</sup> <http://www.vifanord.de> > Sucheinstiege – geographisch.

<sup>13</sup> Der Begriff „Sampo“ stammt aus dem finnischen Nationalepos und bezeichnet ein Gerät, das dem Besitzer durch seine Zauberkräfte Wohlstand verschafft. Er wurde u.a. mit „Zaubermühle“ übersetzt.

<sup>14</sup> <http://www.herder-institut.de/startseite/sammlungen/bilder/ausstellungen/verbreitung-der-ausstellungsorte.html>

<sup>15</sup> <http://www.oclc.org/research/activities/mapfast.html>

„GeoBib – Frühe deutsch- bzw. polnischsprachige Holocaust- und Lagerliteratur – Annotierte und georeferenzierte Online-Bibliographie zur Erforschung von Erinnerungsnarrativen“ ermöglichen, an dem das Herder-Institut neben zahlreichen Arbeitsbereichen der federführenden Universität Giessen beteiligt ist. Im Rahmen dieses bis 2015 laufenden Projekts, für das seitens des Herder-Instituts die Forschungsbibliothek verantwortlich zeichnet, sollen bibliografische Daten - angereichert um Zusatzinformationen zu Personen, Zeiten und Orten - in digitaler Form auf einer Internet-Plattform zugänglich gemacht werden. Ziel ist dabei die Schaffung eines "virtuellen Atlanten", in dem u.a. die wichtigsten Gettos und Lager verzeichnet sein werden. Diese Informationen werden kombiniert mit einem umfassenden Verzeichnis der bis 1949 publizierten Texte, die mit diesen Orten zusammenhängen, ergänzt durch Rezensionen, Sekundärliteratur, Bildquellen sowie - eine wesentliche Besonderheit des Projekts - durch Geodaten. Die Georeferenzierung wird dabei eine Verknüpfung der Werke mit anderen wichtigen, auf die jeweiligen Orte und Regionen bezogenen Informationen erlauben.

Dieses Vorhaben, zwei weitere Kooperationsprojekte des Herder-Instituts, der „Historisch-topographische Atlas schlesischer Städte<sup>16</sup>“ und der „Digitale Atlas politischer Raumbilder zu Ostmitteleuropa<sup>17</sup>“, vor allem aber die im Herder-Institut in Form seiner Kartenabteilung vorhandene Expertise wecken die Hoffnung, dass die im Rahmen dieser Projekte vorhandene Vernetzung wesentlich dazu beitragen wird, die Geodaten-gestützte Präsentation der eigenen Bestände in den kommenden Jahren deutlich vorantreiben zu können.

### **Zusammenfassung und Ausblick**

Aus Platzgründen war es nicht möglich, alle Formen der Vernetzungen von Menschen, Daten und Systemen, an denen die Forschungsbibliothek des Herder-Instituts beteiligt ist, hier ausführlich zu beschreiben. So wurde etwa darauf verzichtet, das *Networking* detaillierter darzustellen, mit dem sich die Forschungsbibliothek als Teil des Herder-Instituts in den Fachcommunities sowohl der Ostmitteleuropaforschung als auch der Bibliotheks- und Informationswissenschaft, in der Lehre an der Universität Giessen, in der Nachwuchsförderung,

---

<sup>16</sup> [http://www.herder-institut.de/staedteatlas\\_schlesien/atlas.html](http://www.herder-institut.de/staedteatlas_schlesien/atlas.html)

<sup>17</sup> <http://www.herder-institut.de/startseite/projekte/laufende/digitaler-atlas-politischer-raumbilder-zu-ostmitteleuropa-im-20-jahrhundert.html>



in der regionalen, nationalen und internationalen Verbandsarbeit sowie in wissenschaftlichen Gesellschaften und Verbänden engagiert. Ebenso wenig konnte die – wiederum auf einer engen Vernetzung mit in- und ausländischen Wissenschaftlern beruhende – wissenschaftliche Tätigkeit von Bibliotheksmitarbeitern gewürdigt werden, die einem der Credos der letzten zwei Evaluierungsberichte des Herder-Instituts entspricht, wonach nur in die Wissenschaft eingebundene Mitarbeiter in der Lage sind, die von Fachwissenschaftlern erwarteten Infrastrukturaufgaben zu leisten. Das Herder-Institut hat dieser Anforderung u.a. durch die Schaffung sogenannter Rotationsstellen<sup>18</sup> entsprochen, die es Mitarbeitern ermöglichen, befristet von ihren eigentlichen Aufgaben entbunden zu werden und sich in dieser Zeit wissenschaftlichen Fragestellungen und Projekten widmen zu können.

Ebenfalls nicht in der eigentlich erfordernden Tiefe eingegangen wurde auf die vergleichsweise geringe Rolle, die elektronische Medien bisher in der Forschungsbibliothek des Herder-Instituts spielen. Dies liegt zum einen am Fehlen einer kritischen Masse kommerzieller Angebote für die historische Ostmitteleuropaforschung, zum anderen aber auch an einer gewissen Zurückhaltung seitens der Klientel der Forschungsbibliothek, die auch die bereits vorhandenen Angebote nur in vergleichsweise geringem Maße nutzt.

Eine gewisse Änderung könnten in Zukunft nicht zuletzt die Angebote des Fachrepositoriums OstDok auslösen, insbesondere dann, wenn es zu der von den Projektpartnern wie auch den Geldgebern gewünschten Veränderung im Publikations- und Rezeptionsverhalten wissenschaftlich anspruchsvoller Texte und zu einer noch weitergehenden Akzeptanz des Open-Access-Gedankens in den Geistes- und insbesondere den Geschichtswissenschaften kommt.

Diese auch politisch gewünschte Stärkung der Prinzipien des Open Access im Rahmen des wissenschaftlichen Publikationsprozesses wie auch die durch das Internet und insbesondere das Social Web auftretenden Veränderungen der Arbeitsgewohnheiten von Wissenschaftlern und Forschern wird sicherlich auch die historische Ostmitteleuropaforschung erreichen und vor neue Herausforderungen stellen.

---

<sup>18</sup> Derartige Rotationsstellen wurden am Herder-Institut erstmals im Rahmen des Schwerpunkts „Kulturtechniken und ihre Medialisierung“ der hessischen Landesexzellenz-Initiative LOEWE geschaffen und dabei auch durch Bibliotheksmitarbeiter wahrgenommen. Nach Auslaufen des LOEWE-Schwerpunkts wurde eine analog konzipierte Rotationsstelle in den Kernhaushalt des Instituts übernommen.

Die Frage, wie wissenschaftliche Spezialbibliotheken auf die sich abzeichnenden Veränderungen im Nutzungsverhalten wie im wissenschaftlichen Informations- und Kommunikationsverhalten angemessen reagieren können, ist sicherlich nur schwer zu beantworten. Für die Forschungsbibliothek des Herder-Instituts scheinen jedoch zwei grundlegende Tendenzen klar: Zunächst wird es darum gehen, den traditionell umfassenden und hochspezialisierten Bestand an Printmedien, für den in den Jahren 2012 bis 2014 durch eine Erweiterungsbaumaßnahme erheblicher zusätzlicher Platz gewonnen wird, in jeder Hinsicht weiterzuentwickeln und auf dieser Basis eine immer stärkere Akzentuierung der Bibliothek als Forschungsbibliothek mit individueller Servicequalität und einem – auch dank der überaus ansprechenden Lage auf dem Marburger Schlossberg – exklusiven Charakter voranzutreiben. Ebenso wird die Forschungsbibliothek aber ihr Angebot elektronischer Medien gezielt ausbauen und in Zusammenarbeit mit den übrigen Sammlungen des Hauses sowie mit eingeführten Partnern von außerhalb auch dadurch ihre Präsenz im virtuellen Raum weiter stärken. Voraussetzung für das eine wie für das andere ist eine funktionierende Vernetzung – von Menschen, Daten und Systemen, wie sie für die Bibliothek seit über sechzig Jahren charakteristisch ist.

### **Literaturverzeichnis**

Die Bibliothek des Herder-Instituts : Geschichte – Bestände – Benutzung. Marburg 1998.

Empfehlungen des Wissenschaftsrats zum Ausbau der wissenschaftlichen Einrichtungen. Bd. II: Wissenschaftliche Bibliotheken, o.O. 1964, S. 224 ff.

Die historischen Reichsgebiete und die Siedlungsgebiete der Deutschen in Ostmittel-, Ost- und Südosteuropa in Bibliotheken der Bundesrepublik Deutschland. Bearb. von Horst von Chmielewski. Köln 1994.

Alle in den Fußnoten aufgeführten URLs wurden zuletzt am 16. August 2012 aufgerufen.

---

## Vernetzung von Datenbanken geisteswissenschaftlicher Forschungsprojekte

Frank Förster<sup>1</sup>, Michael Nolde<sup>2</sup>, Bernhard Thalheim<sup>3</sup>

**Christian-Albrechts-Universität zu Kiel**

<sup>1</sup> *Institut für Klassische Altertumskunde*

<sup>2</sup> *Geographisches Institut (Sektion Geographie)*

<sup>3</sup> *Institut für Technologie der Informationssysteme*

### Zusammenfassung

Das Projekt „Daten- und Informationsmanagement als ein Service“ (DIM-asa-Service) nutzt die heterogene Strukturlandschaft von Datenbanken geisteswissenschaftlicher Forschungsprojekte als Grundlage für eine neuartige Datenbankinfrastruktur. Diese Datenbanken sind in der Regel auf eine spezifische Nutzung ohne vernetzte Bezugspunkte (Insellösung) hin zugeschnitten und die gewonnenen Daten stehen zudem nur einem eingegrenzten Personenkreis zur Verfügung. Die gespeicherten und prozessierten Daten zeichnen sich zudem durch unterschiedliche Abstraktion, Granularität und Qualität in Abhängigkeit von der Forschungsfrage aus.

Ein zentraler Bestandteil bei DIM-asa-Service ist eine Plattform in der Art einer virtuellen Arbeits- und Forschungsumgebung, mit der Daten und Informationen archäologischer, historischer oder kulturwissenschaftlicher Provenienz integriert und genutzt werden können. Neben der Verknüpfung mit Datenbanken beteiligter Projektgruppen wird auch der anreichernde Verweis auf Digitale Bibliotheken und Normdateien möglich. In der Regel besitzen die Projektdaten einen räumlichen Bezug, weshalb ein GIS-Webserver-basiertes Geoportal als Kernkomponente des DIM-asa-Service aufgebaut wird. Mit seiner Hilfe wird die ortsbezogene Suche und die Visualisierung der verfügbaren Daten ermöglicht.

### Abstract

The project “Data and Information Management as a Service” (DIM-asa-Service) uses the heterogeneity of humanist research databases as basis for a novel database technique. Those databases are tailored to a specific use without linked points of reference (isolated applications) and the data gained within research projects are only available for a limited group of people. The stored and processed data are characterized by different abstraction, granularity, and quality, dependent on the research question.

Central tool of DIM-asa-Service is a platform in the manner specified by virtual work and research environments, integrating and using data and from disciplines like archaeology, history or cultural studies. Apart from the linkage with databases of participating research groups, it becomes possible to enrich the database by information from digital libraries and authority files. Generally, the research data have

a spatial reference; therefore, a core component of DIM-asa-Service is a geo-portal based on a GIS webserver. A location-dependent search and visualisation of available data will be enabled with its help.

### Einleitung

In der geisteswissenschaftlichen Forschung werden mehr noch als in vielen Naturwissenschaften Daten meist lokal durch den einzelnen Benutzer in dessen Arbeitsumgebung ohne ausreichende professionelle Informatikerbegleitung verwaltet. Dies führt zu vielen Datenbank-Lösungen, die meist nicht optimal sind, die kaum integriert und von anderen Partnern benutzt werden können und die aufgrund der Projektdauer veralten bzw. nach Projektende verweisen. Man kann zwei Lösungen für diese Datenbank-„Ökosysteme“ anstreben:

- (a) eine gesamtheitliche und zentrale Verwaltung aller Datenbestände mit entsprechenden Diensten (*local-as-view*: lokale Lösung als Sicht auf das globale System);
- (b) eine zentrale Speicherung mit lokaler Nutzung durch die Benutzer, wobei das globale System als Abonnent der lokalen Systeme agiert und die lokalen Systeme eine Dienstleistung des globalen Systems für die eigenen Daten und ggf. auch die Partnerdaten erhalten.

Lösung (a) entspricht dem *Cold-Turkey*-Zugang bei der Migration von Systemen. Lösung (b) dagegen entspricht dem *Butterfly*-Zugang. In der Historie von Migrationsprojekten ist Lösung (a) bis auf eine Handvoll von Anwendungen stets gescheitert. Umgekehrt ist die Erfolgsrate von (b), weshalb wir diesen Zugang wählen. Wir verfeinern ihn durch eine Entwicklung eines leicht zu integrierenden Dienstes.

### Forschungsstand, Defizite

Das „Daten- und Informationsmanagement für geisteswissenschaftliche Forschung als ein Service“ (DIM-asa-Service), das an der Universität Kiel entsteht, nimmt die heterogene und exklusive Natur universitärer Forschung in den Geisteswissenschaften als Grundlage für eine neuartige Datenbankinfrastruktur.

Mit dem Hintergrundwissen um aktuelle Entwicklungen in den Institutionen, die die Bewahrung des kulturellen Erbes pflegen und in die Web-Umgebung der Deutschen Digitalen Bibliothek und Europeana überführen (Bibliotheken, Museen und Archive), soll ein Bereich geisteswissenschaftlicher Forschung, vor allem in philologisch, historisch, kulturwissenschaftlich und archäologisch arbeitenden Disziplinen, bedacht werden, der bisher kaum Forschungszuspruch gefunden hat: Es handelt sich um die wissenschaftlereigene Forschungsdatenbank, welche die nach individuellem Ermessen und Interessen entlang einer Forschungsfrage aggregierten, ausgewerteten und angereicherten Daten und Informationen enthält.

#### Unterschied zu datenintensiver Forschung

Diese Aggregation, Auswertung und Anreicherung geschieht in zunehmendem Maße automatisiert. Die genannten Erinnerungsinstitutionen vermehren das Netz an Vorgaben und Angeboten: zum Beispiel mit Metadaten angereicherte Digitalisate und die nach Standards orientierte Ausarbeitung von Normdateien<sup>1</sup>, aber auch Open Access, Publikationsserver, Repositorien und Virtuelle Forschungsumgebungen.

Zentral sind hierbei die Forschungsdaten. Sie sind als eines der Handlungsfelder im Gesamtkonzept der Informationsinfrastruktur für Deutschland definiert.<sup>2</sup> Eine klare Verantwortlichkeit für die Sammlung, Erschließung, Bereitstellung und Archivierung in Forschungsdatenzentren wird angestrebt, wobei vor allem datenintensive Forschungsaktivitäten im Fokus stehen. Verschiedenste Services für *Digital Science* sind gefragt: Digitale Archive, Bibliothek 2.0-Anwendungen, Tools für die Suche, für Kooperationen und für die Integration und Verteilung von Daten, (vgl. Hashimoto et al. 2012, S. 391–396). Die wissenschaftliche Fachgemeinschaft profitiert davon, da neue Forschungsfragen auf der so erhobenen, empirischen Basis gründen.

---

<sup>1</sup> „Erst die Erschließung von Digitalisaten durch strukturierte, nach internationalen Standards erstellten Metadaten ermöglicht gezieltes Auffinden, Nachnutzen und Interpretieren von Digitalisaten und deren Langzeitarchivierung. Strukturdaten, Normdaten und die Anreicherung mit weiteren Kontextinformationen ermöglichen die umfassende Verknüpfung eines digitalen Objekts mit anderen Objekten („contextual linking“), schaffen so Mehrwerte und lassen Digitalisate häufig erst für einen größeren Kreis von Wissenschaftlern nutzbar werden. Das Potenzial derartiger Verknüpfungen kann nur dann ausgeschöpft werden, wenn möglichst viele Metadaten und im Volltext vorliegende Dokumente frei verfügbar und beliebig miteinander kombinierbar sind. Durch die Weiterentwicklung automatisierter Erschließungs- und Indexierungsverfahren unter Einbezug von Normdateien sollten daher die Möglichkeiten zu einer (halb)automatisierten Anreicherung von Digitalisaten gezielt ausgebaut werden.“ (Fournier 2012, S. 18)

<sup>2</sup> <http://www.leibniz-gemeinschaft.de/?nid=infrastr&nidap=&print=0> (S. 29, 43f., B109f.)

Hingegen basierte geisteswissenschaftliche Arbeit vergleichsweise seltener auf datenintensiver Nutzbarmachung. Ein frühes Beispiel für diese empiriebasierte Form von *Digital Humanities* ist die Worthäufigkeitsanalyse von Pagel, Atkinson und Meade (2007). Publikationen in den Geisteswissenschaften (zumeist Monographien) verlieren jedoch an Substanz, wenn sie nurmehr Auswertungsschrift beigefügter Tabellenwerke sind, weil den Datenvorräten die wesentliche Aussage zugesprochen würde, wie es in naturwissenschaftlichen Aufsätzen oft der Fall zu sein scheint. Dort wird der Wissenschaftler zu einer Art Gehilfen und Dateninterpreten computer- und gerätegenerierter Messgrößen. Datenerhebung und -auswertung werden zum Hauptzweck von Veröffentlichungen. Das Verlassen auf das Anfügen der zugrundeliegenden Daten wird zunehmend gefordert und gefördert, vor allem in den Naturwissenschaften, mit eigens kreierten naturwissenschaftlichen Journals (z.B. Earth System Science Data) und Repositorien (z.B. Pangaea®). Die Grenzen einer geisteswissenschaftlichen Statistik werden dagegen schnell offenbar, denn Sinnverstehen und Blickwinkel geschehen hierin über das Wort, nicht über die Zahl.

Die wesentlichen Unterschiede geisteswissenschaftlicher Datenbanken neugiergetriebener Einzelforscher, die mit dem Begriff *pixel database* beschrieben werden, gegenüber naturwissenschaftlichen Datenbanken liegen

- (a) im Umfang der Daten (relativ gering verglichen mit den dortigen Daten),
- (b) in der Heterogenität der Daten (jedes Projekt mit eigenem Fokus und damit auch eigener Struktur und Auffassung) und
- (c) in einer hohen Strukturdynamik (Auffassungen, welche Daten erforderlich sind und erhoben werden etc., ändern sich ständig, selbst im kleinen Team oder auch beim einzelnen Benutzer).

In einer interdisziplinär verschränkten geisteswissenschaftlichen Forschung sind originäre Forschungsergebnisse von erheblicher Relevanz, wenn sie nicht *primär* auf einer umfangreichen Datenerfassung fußen, sondern aus Textdokumenten, Bohrkernen oder Artefakten eine „Wirklichkeit“ rekonstruieren; dabei ist ein eigens gewählter Anschnitt bibliographischer, objektbezogener und anderer Informationen von Bedeutung, während geisteswissenschaftliche Methoden Pate stehen (vgl. Förster, Thalheim 2011).

#### Unterschied zu Bestrebungen der Erinnerungsinstitutionen

Ein anderer Unterschied geisteswissenschaftlicher Daten thematisch einzeln forschender Wissenschaftler gegenüber aggregierten Daten von Erinnerungsinstitutionen (Metadaten) ist, dass erstere ihre Informationen auf den Punkt hin in meist einfacher Form strukturieren, um ihrer Forschungsfrage Genüge tun zu können, aber für den Augenblick nicht darüber hinaus denken. Im weiteren Verlauf wird Dynamik offenbar, wenn der Forschende die zu Grunde gelegte Untersuchungsfrage rekonzipiert und damit die Daten um weitere Bezüge und (Hintergrund-)Informationen erweitert. Eine unstimmig wirkende Datenstrukturlandschaft legt von dieser inhärenten Dynamik Zeugnis ab.

Metadaten aus Bibliothekskatalogen und Museums- und Archivinventaren liefern wertvolle Hintergrundinformation als Anreicherung der eigenen Daten. Vorteil hierbei ist, dass Erinnerungsinstitutionen eine umfassende Dokumentation der eigenen Bestände in Hinsicht auf eine adäquate Findbarkeit einer nicht klar festzumachenden, quasi entpersonalisierten Nutzergemeinschaft versprechen (Weilenmann 2012). Besonderes Vertrauen genießen hierbei zu Recht Normdateien. Im internationalen Maßstab und deswegen besonders elaboriert liegt die Normdatei für Personen (Virtual International Authority File<sup>3</sup>) vor. Im deutschsprachigen Raum ist die Gemeinsame Normdatei (GND) der Deutschen Nationalbibliothek als Standard zu definieren: es können hieraus Metainformationen zu Personen, Körperschaften, Geographica und Sachschlagwörtern bezogen werden.

Als hilfreiche Idee erweist sich die Überführung der benutzten Metadatenformate EAD, MARC, museumdat, METS, MODS, LIDO und Dublin Core in ein einheitliches, „spartenübergreifendes internes Datenformat“, wie es für die Deutsche Digitale Bibliothek geplant ist (Deutsche Digitale Bibliothek 2010, S. 79f.). Aber im Hinblick auf eine individualisierte Vereinfachung gezielt sortierter Forschungsdaten ist dieses Vorgehen hinderlich.

Ergebnis von all dem ist, dass Datenbanken geisteswissenschaftlicher Forschungsprojekte oftmals als Insellösungen organisiert werden müssen; Bibliothekskataloge etc. dagegen mutieren zu standardisierten und reglementierten Verzeichnissen umfangreicher Prägung. Eine adäquate Verknüpfung der entsprechenden Datenfelder in den Katalogen an die jeweiligen Datenfelder in der

---

<sup>3</sup> siehe [www.viaf.org](http://www.viaf.org)



forschungseigenen Datenbank nach den Vorgaben der Normdateien via *Persistent Identifier* lässt zu, dass Brücken zu und zwischen diesen Insellösungen geschlagen werden können. Eine Vergleichbarkeit auf einer geographischen, personengebunden oder anderen Achse wird somit bei thematisch ähnlich gelagerten Forschungsvorhaben möglich (vgl. Förster, Thalheim 2012).

### **Darstellung der Methodik**

Auf mehreren Ebenen unterstützt das Projekt DIM-asa-Service den Aufbau einer individuellen (forscher- wie forschungseigenen), geisteswissenschaftlichen Datenbank. Mit der Entwicklung neuartiger Techniken werden generische, zusammenstellbare und konfigurierbare Dienste Ausgangspunkt für die jeweiligen Datenbestände werden. Den damit verbundenen Herausforderungen nimmt sich das Projekt an, wobei es neben der vorhandenen bibliothekswissenschaftlichen Expertise auf zwei fachlichen Beinen steht: Informatik und Geographie.

- Informatik: Generische Datenbanken, inklusive einem systematischen Daten- und Informationsmanagement, generischen Integrations- und Publikationsdiensten, einer Unterstützung der Suche und Kooperationsdiensten
- Geographie: Geoportal, inklusive Zeit- und Raumdiensten

Die Graduiertenschule „Human Development in Landscapes“ an der Christian-Albrechts-Universität zu Kiel stellt Forschungsprojekte zur Verfügung, die als Versuchsfolie für das Projekt dienlich sind. Im Mittelpunkt steht die Entwicklung des Menschen und menschlicher Gesellschaften innerhalb von Erdräumen und Landschaften. „Landschaft“ wird hierbei als dynamischer Raum von sozialer, kultureller und ökologischer Bedeutung erfasst. Ein Beispiel soll zur Anschauung genannt werden: Das Aufspüren der Verteilung prähistorischer Siedlungsstrukturen auf Mallorca dient dem Nachweis, wie und in welchem Maßstab in früheren Zeiten die Baleareninsel besiedelt wurde; beispielhaft wird dies bereits anhand einer Auswahl an Literatur vorgeführt (Marx et al. 2012).

Es wird angestrebt, eine plattformunabhängige Open-Source-Lösung für ein lokales Daten- und Informationsmanagement kooperierender Forschergruppen zu erstellen.

### Datenbanktechnik

Objekt-relationale Datenbank-Systeme setzen eine relativ stabile Strukturierung der Daten über einen längeren Zeitraum voraus. Diese ist im Projekt nicht gegeben. Dagegen sind XML-basierte Systeme noch nicht so weit ausgereift, dass sie für relativ dynamische Anwendungen genutzt werden können. Es ist damit möglich, den Daten die Strukturierung auf der Grundlage von RDF-Tripeln mitzugeben. Damit können beliebige Daten abgelegt werden. Deren Nutzung und deren Integration stellt jedoch eine kaum zu bewältigende Herausforderung dar, weil mit jeder Veränderung der Strukturen, mit neuen Datentypen und mit neuer Nutzung eine ständige Erweiterung der Programme vorgenommen werden muss. Dies führt sehr schnell zu chaotischen Systemen mit unlösbaren Wartbarkeitsproblemen. Unser Ansatz sieht dagegen eine automatische Erfassung der Strukturen vor, aus denen die Eingabesichten und auch entsprechende Ausgabesichten automatisch generiert werden können. Damit kann zum einen der volle Umfang für die Speicherung in lokalen Systemen erhalten bleiben. In unserem Ansatz trennen wir die Speicherung von Daten von den Berechnungen über den Daten. Die Berechnung erfolgt weiterhin in den lokalen Systemen. Sie ist deshalb ein überflüssiger Bestandteil von zentralisierten Systemen. Damit bleibt die volle Funktionalität der lokalen Systeme unangetastet, wobei aber gleichzeitig alle Daten global vorgehalten werden können. Hinzu kommt die Technologie der Input- und Outputformen, mit der es möglich ist, Sichten für die Zusammenarbeit mit Partnern über den Daten einfach zu deklarieren und aus diesen Deklarationen dann die erforderlichen Programme automatisch zu erzeugen.

Services des Einbringens, der Suche, des (Mit-)Nutzens und der Präsentation von Daten in geisteswissenschaftliche(n) Datenbanken werden in wiederholbare Teilszenarien, „Ministories“, zerlegt. Die jeweils notwendigen Ministories kann der einzelne Forscher zu einem eigenen Benutzungsszenarium zusammenstellen und so sein eigenes Daten- und Informationsmanagement konfigurieren.

### Geoportal

In der Regel besitzen die Projektdaten einen räumlichen Bezug, weshalb ein GIS-Webserver-basiertes Geoportal als Kernkomponente des DIM-asa-Service aufgebaut wird (Technische Übersichtsskizze: Abb. 1). Mit seiner Hilfe wird die ortsbezogene Suche, die Bereitstellung und die Visualisierung der für die entsprechenden Untersuchungsräume digital und analog verfügbaren Daten ermöglicht.

Sowohl Metadaten als auch Geometriedaten werden hierfür in einer relationalen

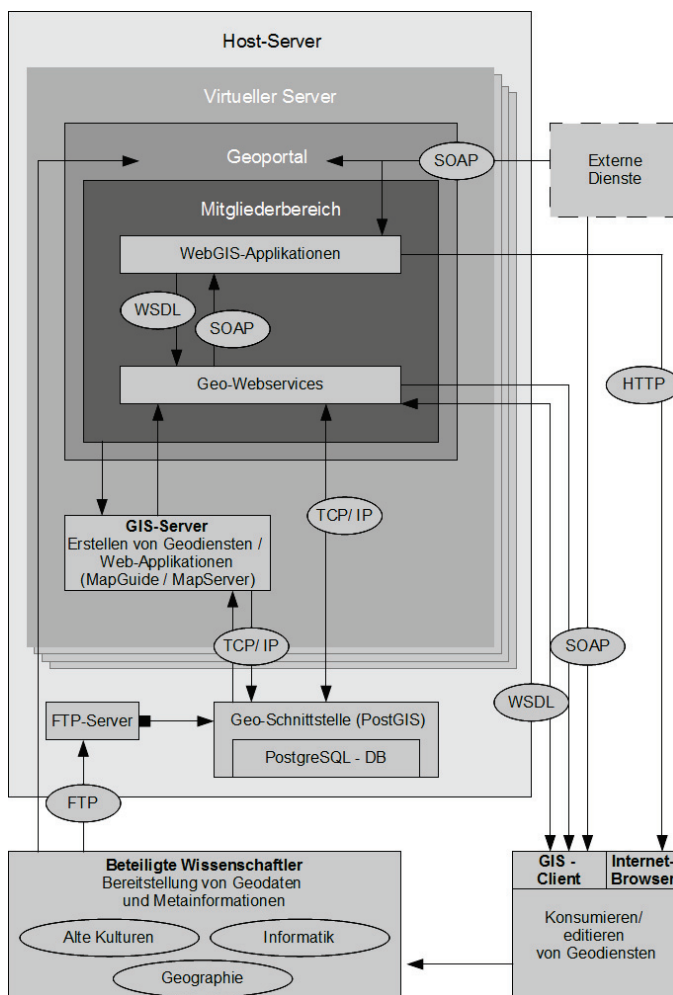


Abb. 1: Technische Übersichtsskizze des DIM-asa-Service-Geoportals

Datenbank abgelegt. Diese ist über spezielle Suchdienste abfragbar, die der Benutzer über das zu erstellende Geoportal steuern kann. So ist nicht nur die Suche nach bestimmten Schlüsselworten, sondern auch die Abfrage eines bestimmten geographischen Gebietes oder eines für die Themenstellung relevanten Zeitraumes möglich. Die Suchergebnisse werden wiederum als Dienst bereitgestellt, sie können direkt im Geoportal in einer integrierten Web-Mapping-Anwendung dargestellt oder in ein Desktop-GIS eingebunden werden.

Solche autark operierenden Such-, Betrachtungs-, Editier- und Downloaddienste sind frei kombinierbar, sodass automatisierte Arbeitsabläufe schnell an geänderte Problemstellungen angepasst werden können (Kossmann, Leymann 2004). Durch diese Kombination von Diensten, auch „Orchestrierung“ genannt, entstehen prozessuale Mehrwertdienste im Sinne einer Serviceorientierten Architektur (SOA).

Neben der Erhöhung der Zugänglichkeit zu wissenschaftlichen Daten hat das Geoportal darüber hinaus die Aufgabe, durch festgelegte Standards für Meta- und Geometriedaten die Heterogenität der eingestellten Daten zu verringern.

Eine feingliedrige Rollen- und Rechteverteilung regelt die Betrachtungs- und Editierrechte der jeweiligen Benutzer.

Bei der Umsetzung wird das Open Source – Web Development Framework „GeoDjango“ verwendet.<sup>4</sup> Dieses ist besonders für die Entwicklung komplexer, datenbankgestützter Anwendungen konzipiert und zeichnet sich durch seinen Funktionsumfang in Hinblick auf eine geographische Fragestellung aus (Ernest, Kaiser 2008). Zudem lässt sich der Funktionsumfang durch Python-Skriptprogrammierung gemäß den benötigten Funktionen erweitern.

### **Diskussion und Interpretation**

Das Projekt DIM-asa-Service unterstützt das traditionelle Rollenverständnis einer individuellen Arbeits- und Forschungsumgebung, wie sie in den Geisteswissenschaften üblich ist. Den Forschenden wird die Kreation einer Datenbank ermöglicht, die an den eigenen Vorgaben einer Forschungsfrage orientiert ist.

---

<sup>4</sup> Projektseite: <http://geodjango.org/>

Die zu Grunde gelegte Datenbanktechnik gestattet als Mehrwert die Verknüpfung der einzelnen Datenbanken im Sinne einer verbesserten Sichtbarkeit der Gesamtergebnisse vernetzter Forschergruppen. Durch die problemlose Ermöglichung dynamischer Erweiterungen ist die Nachnutzbarkeit gewährleistet. Die Zerlegung in Dienste für das Einbringen, das Suchen, das (Mit-)Nutzen und die Präsentation von Daten sichert eine individuelle Zusammenstellung für die eigenen Bedarfe zu.

Zusammengefasst ergibt sich ein Arbeitsprogramm für das Forschungsdatenmanagement individueller geisteswissenschaftlicher Projekte, das

- durch ein einheitliches Daten- und Informationsmanagement von miteinander zusammenarbeitenden Forschern geisteswissenschaftlicher Ausrichtung gekennzeichnet ist,
- eine Vermittlerrolle zwischen universitärer Forschung und dem Sammlungs- und Dokumentationsmanagement von Erinnerungsinstitutionen anstrebt,
- die Probleme zu lösen versucht, die eine wissenschaftsadäquate Integration von Geodateninfrastrukturen derzeit noch hemmen und
- eine langfristige Sicherung und Zugänglichkeit heterogener wissenschaftlicher Daten (analog wie digital) ermöglicht.

Die Idee einer forscher- und forschungsnahen Daten- und Informationsinfrastruktur für die Geisteswissenschaften wird mit DIM-asa-Service trefflich unterstützt.

### Literaturverzeichnis

- (1) Hans-Georg Becker, Frank Förster: *Vernetztes Wissen – Ereignisse in der bibliographischen Dokumentation*. – In: Zeitschrift für Bibliothekswesen und Bibliographie 57 (2010) 1, S. 15–25.
- (2) Deutsche Digitale Bibliothek: Rahmenbedingungen zur Anforderungsanalyse aus politischer, rechtlicher und funktionaler/technischer Sicht. Finale Fassung, 21. Mai 2010. Hg. vom Fraunhofer IAIS.
- (3) Johannes Ernesti, Peter Kaiser: Python 3. Bonn: Galileo Press, 3., aktualisierte und erw. Aufl. 2008.
- (4) Frank Förster, Bernhard Thalheim: *Data and Information Management for Humanist Researchers (Text and Things)*. Supporting Digital Humanities 2011, University of Copenhagen, 17./18.11.2011.
- (5) Frank Förster, Bernhard Thalheim: *An effectual approach for a data and information management for humanists*. 4th International Conference on Qualitative and Quantitative Methods in Libraries, Limerick, 22.–25.5.2012.
- (6) Frank Förster: *Integrating Knowledge from Libraries, Museums etc. into University Culture*. Knowledge / Culture / Social Change International Conference, University of Western Sydney, 7.–9.11.2011.
- (7) Johannes Fournier: *Zugang, Nachnutzung und Reproduzierbarkeit. Anmerkungen zur künftigen Ausrichtung einer wissenschaftsadäquaten Informationsinfrastruktur*. – In: Bibliothek, Forschung und Praxis 36 (2012). Preprint.
- (8) Takako Hashimoto, Jaak Henno, Hannu Jaakkola, Ana Šaša, Bernhard Thalheim: *Infrastructures for Knowledge Systems Environments*. – In: Information Modelling and Knowledge Bases XXIII (Frontiers in Artificial Intelligence, Bd. 237). Amsterdam: IOS Press, 2012, S. 369–398.
- (9) Kai Jannaschk, Claas Anders Rathje, Bernhard Thalheim, Frank Förster: *A Generic Database Schema for CIDOC-CRM Data Management*. – In: ADBIS 2011 Research Communications. Proceedings II of the 15th East-European Conference on Advances in Databases and Information Systems, Vienna, Austria, September 20-23, 2011. Hg. von Johann Eder, Maria Bielikova und A Min Tjoa, S. 127–136.
- (10) Donald Kossmann, Frank Leymann: *Web Services*. – In: Informatik-Spektrum 27 (2004) 2, S. 117–128.
- (11) Kirstin Marx, Christoph Rinne, Monica de Cet, Rainer Duttmann, Rolf Gabler-Mieck, Wolfgang Hamer, Corinna Kortemeier, Johannes Müller: *Geospatial Analysis of Prehistoric Settlement Structures Considering Monumental Architecture on Mallorca*. – In: „As time goes by?“ Monumentality, Landscapes and the Temporal Perspective. Hg. von Martin Furholt, Martin Hinz und Doris Mischka. Bonn: R. Habelt, 2012, S. 105–112.
- (12) Mark Pagel, Quentin D. Atkinson, Andrew Meade: *Frequency of word-use predicts rates of lexical evolution throughout Indo-European history*. – In: Nature 449 (2007), S. 717–720 [doi:10.1038/nature06176].
- (13) Anne-Katharina Weilenmann: *24/7/365 x 360 – die neue Zauberformel?* Conference Proceedings Paper, WissKom 2012, Jülich.

---

## **Impact Messung – Mehrwertdienste für Repositorien**

Daniel Beucke

Niedersächsische Staats- und Universitätsbibliothek, Göttingen

### **Zusammenfassung**

Das DFG-geförderte Projekt Open-Access-Statistik sammelt und aggregiert Nutzungsdaten von Open-Access-Repositorien und erstellt darauf basierende international vergleichbare Nutzungsstatistiken. Nach Projektende im Frühjahr 2013 wird der Dienst an die Verbundzentrale des GBV übergeben und dort dauerhaft betrieben. Im Gegensatz zu zitationsbasierten Metriken haben Nutzungsstatistiken digitaler Dokumente den Vorteil der zeitnahen Messung und ermöglichen somit aktuelle Angaben zur Relevanz. Eine weitere neue Entwicklung sind alternative Metriken, die in der Fachcommunity diskutiert werden. Diese basieren auf Ereignissen aus Social-Media-Diensten wie Bookmarks oder Kommentaren.

### **Abstract**

The DFG-funded project Open Access Statistics collects and aggregates usage data from open access repositories and offers internationally comparable usage statistics. After the end of the project in spring 2013, the service will be offer by the Federation headquarters of GBV permanently. In contrast to citation-based metrics are the usage statistics of a digital document a nearly realtime measurement of its relevance. Altmetrics are a new development in this field. They are based on events from social media services such as bookmarks or comments.

### **1. Motivation**

Die Prozesse wissenschaftlicher Kommunikation, ihre Medien und Akteure unterliegen einem stetigen Wandel. Durch die Erfindung der beweglichen Lettern durch Gutenberg im 15. Jahrhundert wurde die Buchherstellung vereinfacht und damit erschwinglich. Diese Medienrevolution kann als Vereinfachung und Demokratisierung der Kommunikation bezeichnet werden. Wissen wurde einer breiteren Masse zugänglich und damit überprüfbar. In den folgenden Jahrhunderten weitete sich diese Form der Kommunikation und Wissensvermittlung immer weiter aus. Zudem wurde die wissenschaftliche Kommunikation durch die Weiterentwicklung von Verkehrsmitteln, Post und Telefon weiter vereinfacht. Diese Kommunikationsformen werden der analogen Kommunikation zugerechnet.



Heute befindet sich die Kommunikation wieder an einem revolutionären Wendepunkt. Mit dem Aufkommen des World Wide Web entwickelte sich eine neue Form der Kommunikation, die auch in der Wissenschaft immer mehr Anwendungsfelder findet: die digitale Wissenschaftskommunikation. "Readers are faced with a 'river of knowledge' that allows them to select among a multitude of sources, and to find near substitutes when necessary" (Odlyzko 2002). Diese Informationsflut gilt es in nächster Zukunft, durch Anwendungen und Dienste zu kanalisieren und nach relevanten Informationen zu filtern.

Ein möglicher Weg für das Filtern der großen Masse an Information ist, sich auf Maßzahlen – sogenannte Impact-Werte – zu stützen. In der Wissenschaft haben sich für die Messung des Impacts verschiedene Mechanismen etabliert. Der sicher bekannteste Messwert ist der Journal Impact Factor (JIF), der auf Grundlage von Zitationen gemessen wird. Der JIF bezieht sich auf die durchschnittliche Anzahl von Zitationen einer gesamten Zeitschrift. Dabei wird nicht betrachtet, wie oft einzelne Artikel, sondern wie oft alle Artikel einer Zeitschrift in einer bestimmten Zeitspanne nach Erscheinen der Zeitschrift im Durchschnitt zitiert wurden. Zur Auswertung werden Datenbanken herangezogen, die von kommerziellen Anbietern gegen Lizenzgebühren zur Verfügung gestellt werden. Seit längerem steht der JIF in der Kritik (vgl. Dong 2005; Seglen 1997) und auch aktuell wird über seine sinnhafte Anwendung diskutiert (vgl. Curry 2012).

Eine weitere Möglichkeit der Impact-Messung ist der Fokus auf den Autor. Dabei erhält der Autor eine Maßzahl, die seinen Impact aufzeigt. Das diesbezüglich wohl bekannteste bibliografische Maß ist der nach seinem Entwickler Jorge E. Hirsch benannte Hirschfaktor (oder h-Index). Es basiert auch auf Grundlage von Zitationen (vgl. Hirsch 2005). Um den h-Index zu berechnen, wird die Anzahl der Publikationen einer Person mit der jeweiligen Anzahl der Zitierungen in Relation gesetzt. So hat beispielsweise ein Autor den h-Index von 7, wenn sieben Publikationen jeweils mindestens siebenmal zitiert wurden.

Eine alternative und ergänzende Messung des wissenschaftlichen Impacts, die nicht auf der Auswertung von Zitationen beruht, ist die Auswertung der Nutzungsdaten von elektronischen Dokumenten. Diese Statistiken bilden das über Downloads gemessene Interesse an einer Veröffentlichung ab. Sie ermöglichen so eine

nutzungsbasierte und von Zitationen unabhängige Bewertung. Nutzungsstatistiken digitaler Objekte können in Echtzeit und mit geringem Aufwand erstellt und unabhängig von ihren Objekteigenschaften wie Publikationsplattform, Materialart, Herkunftsland, Sprache und Themengebiet verglichen werden.

## **2. Das Projekt Open-Access-Statistik**

Diese Vorteile greift das seit 2008 DFG-geförderte Projekt Open-Access-Statistik (OA-Statistik)<sup>1</sup> auf. Ein Ziel ist die Berechnung und Bereitstellung international vergleichbarer und standardisierter Nutzungsstatistiken für wissenschaftliche Open-Access-Repositorien. In der ersten Förderphase (2008 bis 2010) beteiligten sich die Niedersächsische Staats- und Universitätsbibliothek Göttingen, der Computer und Medien Service der Humboldt-Universität zu Berlin, die Saarländische Universitäts- und Landesbibliothek Saarbrücken sowie die Universitätsbibliothek Stuttgart an dem Projekt und bauten eine Infrastruktur auf, die den Austausch und die anschließende Bereitstellung der berechneten Nutzungsstatistiken ermöglicht. In der zweiten Förderphase (2011 bis 2013) unterstützt die Verbundzentrale des GBV als weiterer Projektpartner das Team.

Das Hauptziel der aktuellen Projektlaufzeit ist die Überführung der Projektergebnisse in einen stabilen Dienst und so die Erhöhung der Nutzbarkeit für Betreiber von Repositorien. Dafür wurde nach Ablauf der Hälfte der Projektlaufzeit eine Machbarkeitsstudie (vgl. OA-Statistik 2012) erstellt, die sich mit Themen der technischen Realisierbarkeit, der Finanzierung des Dienstes und den in Deutschland recht restriktiven Datenschutzvorgaben bei statistischen Auswertungen befasst. Bis Ende der Projektlaufzeit wird die OAS-Infrastruktur um weitere Repositorien erweitert sowie die Performanz der Bereitstellung von standardisierten Nutzungsstatistiken optimiert.

### 2.1. International standardisierte Nutzungsstatistiken

Für den Austausch und die Berechnung von Nutzungsstatistiken ist es von hoher Bedeutung, auf Standards zu setzen. Dadurch wird zum einen die Infrastruktur skalierbar und zum anderen können die berechneten Statistiken international verglichen werden. Zur Erreichung dieser Ziele kooperieren die Projektpartner von OA-Statistik mit verschiedenen internationalen Projekten und Institutionen, die sich

---

<sup>1</sup> <http://www.dini.de/projekte/oa-statistik/>

ebenfalls mit Nutzungsstatistiken beschäftigen. Das Austauschformat für die Nutzungsdaten wurde ursprünglich von Bollen (Bollen, Van de Sompel 2006) spezifiziert. Darauf aufbauend hat OA-Statistik als Teil der Knowledge-Exchange-Arbeitsgruppe Usage Statistics<sup>2</sup> Richtlinien zum standardisierten Austausch von Nutzungsdaten auf europäischer Ebene erarbeitet. Die *Guidelines for the aggregation and exchange of Usage Data* beschreiben ausführlich die Schnittstellen, das Austauschformat und das Vorgehen bei der Aggregation von Nutzungsdaten (Verhaar et al. 2010).

Bei der Aggregation von Nutzungsdaten spielt das Projekt Counting Online Usage of Networked Electronic Resources (COUNTER)<sup>3</sup> eine zentrale Rolle. Es definierte einheitliche Standards zur Nutzungsmessung von Informationen. Hierzu wurden die *COUNTER Codes of Practice*<sup>4</sup> entwickelt, die von OA-Statistik als Standard zur Aggregation der Nutzungsdaten und zur Berechnung der Nutzungsstatistiken eingesetzt werden.

Bei der Berechnung der Nutzungszahlen ist ein wichtiger Aspekt das Herausfiltern von nicht-menschlichen Zugriffen, um Verzerrungen zu verhindern. Dazu wird u.a. eine Robot-Blacklist verwendet, auf der die identifizierten Robots verzeichnet sind.

Eine solche Liste liegt zurzeit nicht vor bzw. wird nicht aktualisiert. Eine internationale Zusammenarbeit bei der Erstellung und Pflege einer solchen Liste hat die Vorteile, dass sie wesentlich umfangreicher sein kann und somit die Vergleichbarkeit von Nutzungsstatistiken steigt. Ein Ziel von OA-Statistik ist es daher, am Ende der Projektlaufzeit eine solche Liste international zu etablieren.

## 2.2. Nachhaltigkeit

Mit Beginn der zweiten Förderphase verfolgen die Projektpartner aus OA-Statistik das Ziel, die erreichten Ergebnisse in einen nachhaltigen Dienst zu überführen. Dies kann nur gelingen, wenn sich alle Projektpartner entsprechend ihrer Fähigkeiten und Möglichkeiten langfristig engagieren. Um dieses Engagement zu dokumentieren und

---

<sup>2</sup> <http://www.knowledge-exchange.info/Default.aspx?ID=365>

<sup>3</sup> <http://www.projectcounter.org>

<sup>4</sup> [http://www.projectcounter.org/code\\_practice.html](http://www.projectcounter.org/code_practice.html)

langfristig zu sichern, haben die Partner eine gemeinsame Absichtserklärung<sup>5</sup> verfasst. Darin verpflichten sich alle Partner, im Rahmen der DINI AG „Elektronisches Publizieren“<sup>6</sup> den Kontakt mit internationalen Projekten und Institutionen zu pflegen und neue Kontakte herzustellen, um die lokale und globale Vernetzung des Dienstes OA-Statistik zu gewährleisten.

Essentiell wichtig für einen stabilen Dienst, ist die datenschutzrechtliche Absicherung des Angebotes. In der juristischen Fachdiskussion herrscht noch Uneinigkeit, ob IP-Adressen, die für die Berechnungen in OA-Statistik von hoher Wichtigkeit sind, als personenbezogene Daten zu werten sind. Auf der Grundlage eines bei der Zentralen Datenschutzstelle der baden-württembergischen Universitäten (ZENDAS)<sup>7</sup> in Auftrag gegebenen Gutachtens (vgl. Stemmer 2011) wurde der gesamte Workflow in OA-Statistik datenschutzrechtlich untersucht und daraufhin durch die Projektpartner technisch angepasst. Die Lösung ist die Anonymisierung der IP-Adressen beim verarbeitenden OA-Statistik Service-Provider. Hierzu wurde ein Verfahren gewählt, das technisch die IP-Adressen so verfremdet, dass sie keinerlei Rückschlüsse auf ihre Herkunft möglich sind. Die Anforderungen der ZENDAS werden bis Projektende umgesetzt, sodass der Dienst nach Ende der Projektlaufzeit datenschutzrechtlich abgesichert betrieben werden kann.

Einer der wichtigsten Aspekte bei der Nachhaltigkeit ist das Fortbestehen des Dienstes nach Projektende. Der Dienst muss zur Gewährleistung eines dauerhaften Betriebs von einer vertrauenswürdigen und zuverlässigen Institution übernommen werden, die die technischen und organisatorischen Mittel hat, ihn langfristig anbieten zu können. Eine solche Institution sollte wirtschaftlich stabil und in der Community etabliert sein, sowie die notwendige technische und personelle Infrastruktur besitzen. Zudem sollte sie fundierte Kenntnisse im Betrieb ähnlicher Dienste hinsichtlich Geschäftsgang, Datenkonversion und datenschutzkonformer Datenverarbeitung vorweisen können. In der zweiten Förderphase von OA-Statistik konnte – als potenzieller Dienstanbieter nach Projektende – die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) als Projektpartner gewonnen werden.

---

<sup>5</sup> Siehe: Memorandum of Understanding. Anhang der "Machbarkeitsstudie zum nachhaltigen Betrieb des Dienstes Open-Access-Statistik".

<sup>6</sup> [http://www.dini.de/fileadmin/oa-statistik/Machbarkeitsstudie/Anhaenge\\_Machbarkeitstudie\\_OAS.pdf](http://www.dini.de/fileadmin/oa-statistik/Machbarkeitsstudie/Anhaenge_Machbarkeitstudie_OAS.pdf)

<sup>7</sup> <http://www.dini.de/ag/e-pub>

<sup>7</sup> <http://www.zendas.de>

Sie wird der Dienst OA-Statistik im Laufe des nächsten Jahres in ihr Angebot übergehen.

### 2.3. OA-Statistik Workflow

Für das Erfassen, den Austausch, die Aggregation und das Bereitstellen der aufbereiteten Nutzungsstatistiken ist eine technische Infrastruktur notwendig. Die OA-Statistik Infrastruktur besteht aus zwei Kernkomponenten: Auf der Seite des jeweiligen Repositoriums ist der OA-Statistik Data-Provider (OAS-DP). Dieser muss durch den Repositoriumsbetreiber bereitgestellt werden. Zur Installation kann die Software auf den Webseiten von OA-Statistik heruntergeladen werden. Die Software muss an die spezifischen Anforderungen des eigenen Repositoriums angepasst werden und ist danach einsatzbereit. Auf der anderen Seite ist der OA-Statistik Service-Provider (OAS-SP), der zentral beim Dienstanbieter läuft. Im Folgenden wird der OA-Statistik Workflow (Abb. 1) beschrieben.

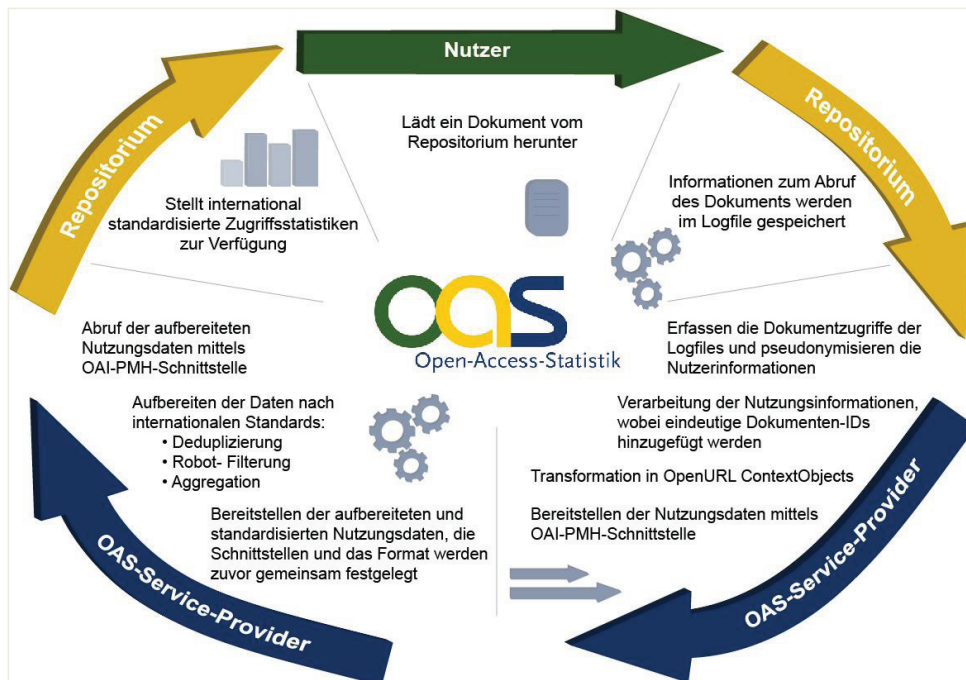
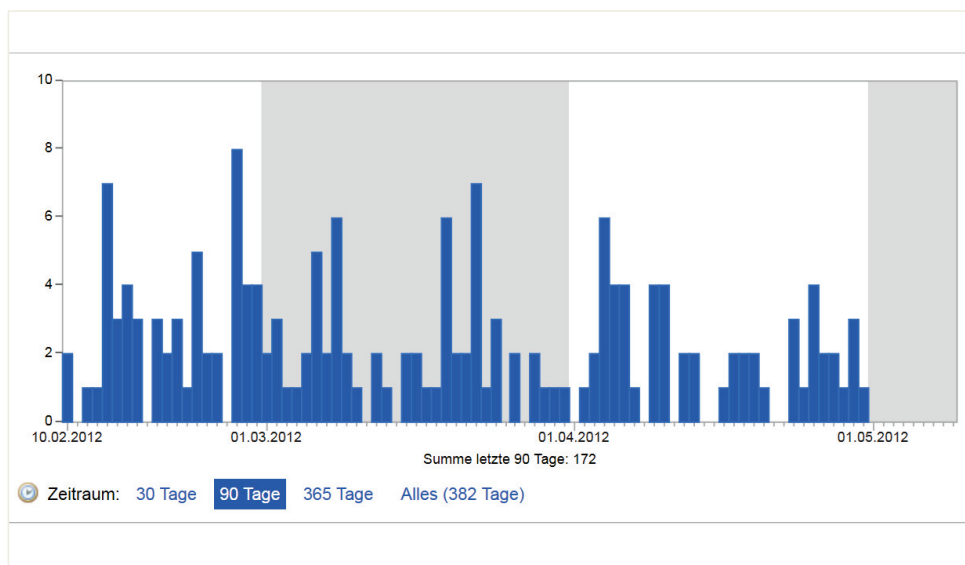


Abbildung 1: OA-Statistik Workflow

Wenn ein Nutzer ein Dokument auf einem Repositorium herunterlädt, so überträgt sein Browser Informationen an den Webserver. Diese Information zum Abruf des Dokumentes werden im Logfile des Repositoriums gespeichert.

Im nächsten Schritt erfasst der OAS-DP die Dokumentzugriffe der Logifiles und bereitet sie nach Vorgaben von OA-Statistik auf. Dabei werden die Nutzerinformationen wie IP-Adresse und Browserkennung pseudonymisiert und eine eindeutige Dokument-ID hinzugefügt. Die so erstellten Daten werden in ein XML-Format (OpenURL ContextObjects) überführt und mittels einer passwortgeschützten OAI-Schnittstelle dem OAS-SP bereitgestellt.

Der OAS-SP ruft die aufbereiteten Nutzungsdaten mittels OAI-PMH ab und bereitet diese nach internationalen Standards auf. Dabei erfolgen die Deduplizierung, die Robot-Filterung und die Aggregation der Zahlen. Abschließend werden die aufbereiteten und standardisierten Nutzungsdaten den Repositorien über Schnittstellen bereitgestellt. Dies erfolgt – nach Absprache mit den Repositorien – in Form von XML, JSON oder als grafische Darstellung (vgl. Abb. 2).



**Abbildung 2: Grafisch aufbereitete Nutzungsstatistiken**

Das Repository kann somit seinen Nutzern Zugriffsstatistiken für die digitalen Dokumente zur Verfügung stellen. Auf den Übersichtsseiten zu den Dokumenten können die Statistiken dargestellt werden, gleichzeitig kann der Repositorysbetreiber die Daten für interne Controllingzwecke nutzen. Darüber hinaus sind weitere Szenarien der Datennutzung denkbar, wie beispielsweise die Sortierung nach Downloadzahlen bei einer Suchanfrage.

### 3. Ausblick

Wie eingangs beschrieben, befindet sich die wissenschaftliche Kommunikation im Wandel. Wissenschaftler verlegen ihre Arbeit immer weiter ins World Wide Web, arbeiten kollaborativ an Publikationen und veröffentlichen ihre Ergebnisse unter anderem in Weblogs. Dadurch bedingt verändern sich die Werkzeuge mit denen sie täglich recherchieren und publizieren. Bei Literaturverwaltungsprogrammen wie Zotero und Mendeley steigen die Nutzerzahlen, sie bieten mittlerweile eine Vielzahl von Artikeln an. Viele Wissenschaftler teilen mittlerweile ihre Erkenntnisse und Ergebnisse über Twitter mit (vgl. Barr 2012).

In der Wissenschaft wird es immer ein Maß geben, mit dem der Output eines Wissenschaftlers gemessen wird. Die zuvor beschriebenen „klassischen“ Impact-Maße auf Grundlage von Zitationen oder die nutzerbasierten Metriken können alleine betrachtet nicht zielführend sein. Vielmehr muss sich ein solches Maß aus verschiedenen Faktoren zusammensetzen. Bollen et al. kamen 2009 in ihrem Artikel *A principal component analysis of 39 scientific impact measures* zu dem Schluss: „Our results indicate that the notion of scientific impact is a multi-dimensional construct that cannot be adequately measured by any single indicator, although some measures are more suitable than others. The commonly used citation Impact Factor is not positioned at the core of this construct, but at its periphery, and should thus be used with caution.“ (Bollen et al. 2009).

Zukünftige Herausforderungen werden sein, die von Bollen et al. genannte Vielfalt von Metriken und die aktuellen Entwicklungen im Zusammenhang mit den Social-Media-Diensten zu verknüpfen. Nur diese Verbindung kann einen Aufschluss über den Impact eines digitalen Objekts geben.

Viele Social-Media-Dienste bieten Schnittstellen, über die Bookmarks, Links oder Diskussionen abgefragt werden können. Die ersten Anbieter machen von diesen Schnittstellen Gebrauch und aggregieren die Daten, um sie aufbereitet ihren Nutzern zur Verfügung zu stellen. Der Open-Access-Verlag PLoS ONE<sup>8</sup> nutzt diese Möglichkeit schon länger und bietet für seine Artikel verschiedene alternative Metriken an (vgl. Abb. 3).

---

<sup>8</sup> <http://www.plosone.org>

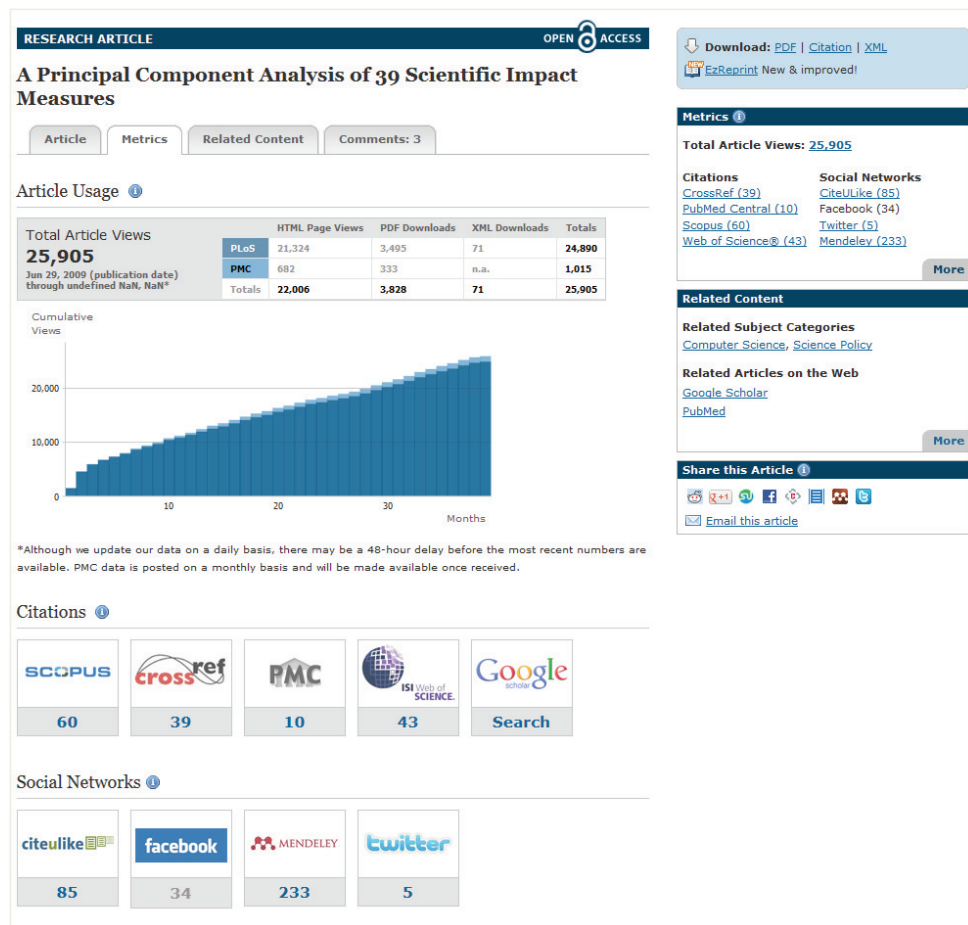


Abbildung 3: Darstellung der Metrics bei PLoS ONE

In Zukunft steigt stetig das Interesse, nicht nur Maßzahlen für Zeitschriften und einzelne Artikel zu erhalten, sondern direkt für den jeweiligen Autor selbst. Auch hier kann auf die Schnittstellen der Social-Media-Dienste zurückgegriffen werden, um solche Daten zu aggregieren und bereit zu stellen. Ein Beispiel ist die Webseite ScienceCard (siehe Abb. 4). Sie aggregiert die Daten nutzerspezifisch und bietet so einen Überblick von Impact-Werten eines Wissenschaftlers.



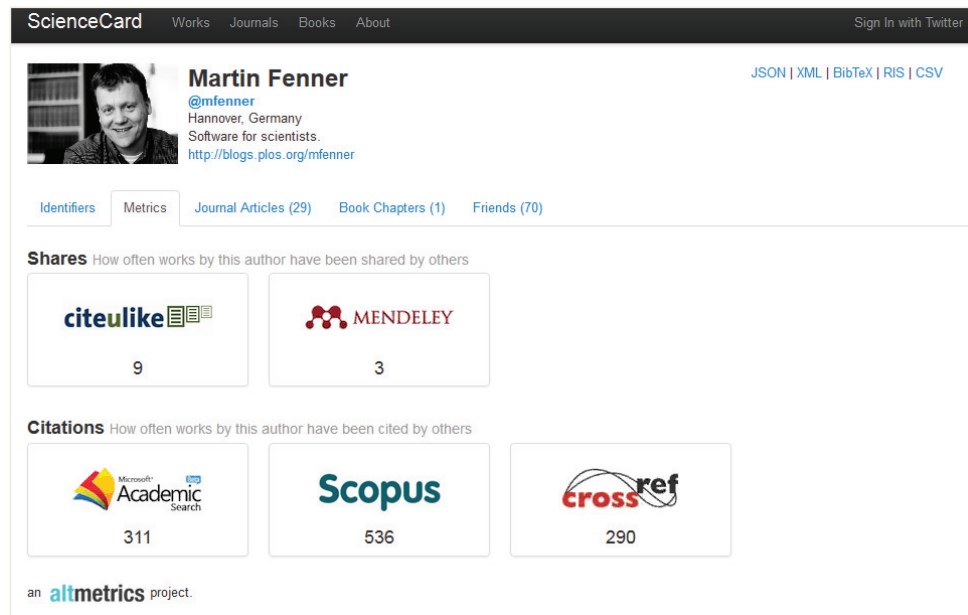


Abbildung 4: Darstellung in ScienceCard

Die Projektpartner aus OA-Statistik verfolgen diese Entwicklungen interessiert und bringen sich in die aktuellen Diskussionen mit ein. Für das folgende Jahr ist ein Projekt geplant, dass diese Entwicklungen aufgreift und auf Repositorien anwendet.

## Literaturverzeichnis

Barr, Kelli (2012): The Role of altmetrics and Peer Review in the Democratization of Knowledge. <http://altmetrics.org/altmetrics12/barr/>.

Bollen, Johan; Van de Sompel, Herbert (2006): An Architecture for the Aggregation and Analysis of Scholarly Usage Data. To be published in Proceedings of the Joint Conference on Digital Libraries. <http://arxiv.org/abs/cs.DL/0605113>.

Bollen, Johan; Van de Sompel, Herbert; Hagberg, Aric; Chute, Ryan (2009): A Principal Component Analysis of 39 Scientific Impact Measures. PLoS ONE 4(6): e6022. doi:10.1371/journal.pone.0006022.

Curry, Stephen (2012): Sick of Impact Factors. <http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/>.

Dong, Peng; Loh, Marie; Mondry, Adrian (2005): The "impact factor" revisited. In: Biomed Digit Libr 2 (1), S. 7. doi:10.1186/1742-5581-2-7.

Hirsch, Jorge E. (2005): An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciences of the United States of America. Bd. 102, Nr. 46, S. 16569–16572. doi:10.1073/pnas.0507655102.

OA-Statistik (2012): Machbarkeitsstudie zum nachhaltigen Betrieb des Dienstes Open-Access-Statistik. <http://www.dini.de/fileadmin/oa-statistik/Machbarkeitsstudie/OAS-Machbarkeitsstudie.pdf>.

Odlyzko, Andrew (2002): The rapid evolution of scholarly communication. In Learned Publishing. Heft 1. 2002 (15. Jg.). S. 7-19.

Seglen, Per O (1997): Why the impact factor of journals should not be used for evaluating research. BMJ. 1997; 314 (7079):498-502.

Stemmer, Bastian (2011): Datenschutzrechtliche Bewertung des Projekts „Open-Access-Statistik“. Zentrale Datenschutzstelle des baden-württembergischen Universität (ZENDAS). [http://www.dini.de/fileadmin/oa-statistik/projektergebnisse/ZENDAS\\_Gutachten\\_2011.pdf](http://www.dini.de/fileadmin/oa-statistik/projektergebnisse/ZENDAS_Gutachten_2011.pdf).

Verhaar, Peter et al. (2010): Guidelines for the aggregation and exchange of Usage Data. Surf Wiki 2010. <http://wiki.surf.nl/display/standards/KE+Usage+Statistics+Guidelines>.

---

## **Spezifische Suchmaschinen mit der Open Source Suchmaschine YaCy am Beispiel der Johannes Gutenberg Universität Mainz**

Jörg Berkemeyer<sup>1</sup>, Georg von Nessler<sup>2</sup>

<sup>1</sup> IHS Global Insight Ltd, Frankfurt am Main

<sup>2</sup> IP-ShareMedia, Frankfurt am Main

### **Zusammenfassung**

YaCy ist eine freie, per Open Source angebotene Suchmaschinentechnik, mit deren Hilfe unstrukturierte Wissensrepositorien indexiert und damit durchsuchbar gemacht werden können. Die Kontrolle der dabei angezeigten Suchergebnisse und Indexierungsverfahren liegt beim Wissensmanager/Bibliothekar und nicht wie bei kommerziellen/proprietären Systemen bei einem kommerziellen Anbieter. Der Wissensmanager bestimmt, was indexiert wird, die Nutzer können die Auswahl der indexierten Quellen überprüfen und ggfs. in Absprache mit dem für die Indexierung verantwortlichen Bibliothekar/Informationsmanager eine Erweiterung der Quellen diskutieren. Das Erschließen von Informationen, die Bereitstellung von Informationen und das Informationsmanagement sind Schlüsselfaktoren, um Innovationen und neues Wissen zu generieren und für die wissenschaftliche Gemeinschaft verfügbar zu machen. Die hierfür aktuell bereitgestellten Instrumente basieren auf Technologien, bei denen für die Nutzer und Informationsmanager die zugrundeliegenden Kriterien der Indexierung unzureichend offen liegen. Hier wird auch die Funktion des Bibliothekars in das Informationszeitalter übertragen, denn sein Fachwissen in Bezug auf die Thematik und seine Kompetenz bei der Auswahl und Indexierung der Inhalte sind eine Schlüsselqualifikation. Der Beitrag stellt die Nutzung von YaCy als individuelle Suchmaschine vor, erläutert die Einsatzmöglichkeiten und vergleicht das Open Source Konzept mit den marktverfügbaren Systemen. Die praktische Anwendung kann anhand der Quellenverweise nachverfolgt werden.

### **Abstract**

Open source search engine technology YaCy makes unstructured knowledge repositories accessible by semantic indexing. Knowledge managers / librarians are in control of the presentation of search results and underlying indexing techniques in contrast to most commercially available or proprietary search engine software. The knowledge manager selects contents that are to be indexed – users are able to see underlying sources and search algorithms enabling both parties to discuss enhancements and adaptations. The classification of information, the dissemination of information and information management are key factors to promote innovation and generate knowledge within the scientific community. The use of proprietary software to solve these tasks limits control and transparency of underlying indexing procedures to both users and information managers. To select adequate tools and software as well as the knowledge in indexing scientific sources are key qualifications of the librarian in the information age. This essay introduces YaCy as a customizable search engine for libraries and compares the open source concept with proprietary systems. YaCy in practice can be experienced via the listed links.

### Einsatz von YaCy an der Universität Mainz

Nachdem die bisher eingesetzte Software Google Search Appliance für die Campusweite Suche an der Johannes-Gutenberg-Universität Mainz bereits zum zweiten Mal ausgefallen war, entschlossen sich die Verantwortlichen der Bibliothek für die Suche zusätzlich auf die OpenSource Software YaCy zu setzen. Als weiterer Grund, der gegen den alleinigen Einsatz der Google Search Appliance sprach, war die Beschränkung bei der eingesetzten Version auf die Indexierung von max. 50.000 Dokumenten. Der Einsatz von YaCy versprach:

- ⌚ Eine höherer Ausfallsicherheit
- ⌚ Eine höhere Skalierbarkeit
- ⌚ Keine Beschränkung bei der Anzahl der zu indizierenden Dokumente

Die Umstellung konnte kurzfristig nach dem zweiten Ausfall von Google Search Appliances erfolgen, seit dem 8.12.2011 werden alle Dokumente mit YaCy indexiert und bereitgestellt.<sup>1,2</sup>

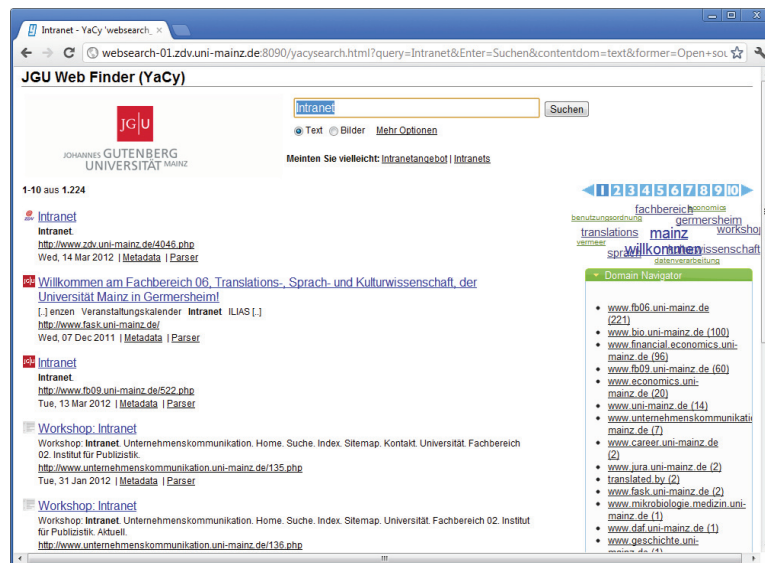


Abb.1 Suchergebnisse der auf interne Inhalte der Universität beschränkten Suche

<sup>1</sup> YaCy ersetzt Google Search Appliance, <http://www.zdv.uni-mainz.de/uni-intern/ot/1131.php> (14.12.2011)

<sup>2</sup> YaCy Suchmaschine der Universität - JGU Web Finder, <http://www.zdv.uni-mainz.de/uni-intern/ot/1131.php> (15.8.2012)

## Die Suchmaschinen-Software YaCy

YaCy ist eine quelloffene Suchmaschinensoftware mit der selbst-definierte Intranet und Extranet-Suchen realisiert werden können. Die Initiatoren von YaCy gehen davon aus, dass

*Die Informationsgesellschaft des 21. Jahrhunderts basiert darauf, dass der Zugang zu allen öffentlichen Informationen frei ist. Das bedeutet: der Zugang sollte transparent, nachvollziehbar und für jeden auch individuell gangbar sein.*<sup>3</sup>

Als Quellen für die YaCy-Suche können ausschließlich interne Dokumente dienen, es sind aber auch externe Dokumentensammlungen in die Suche integrierbar. So wird bereits durch Festlegen der zu durchsuchenden Repositorien Themenrelevanz hergestellt.

YaCy ist ein Kombination aus WebCrawler und Suchmaschine, die als quelloffene Software auf den eigenen Servern betrieben werden kann, die aber auch die Indexierung anderer offen verfügbarer Quellen nutzt. Auf diese Weise lassen sich thematische Suchen definieren, ohne die angezeigten Ergebnisse zu sehr einzuschränken. Die Volltextsuche erfolgt beispielsweise über eine ganze Domain oder ein komplettes Dokumentenverzeichnis. Werbung, irrelevante Inhalte und künstlich für Suchmaschinen optimierte Seiten werden aber nicht berücksichtigt.

Für den Anwender bietet sich eine gewohnte Volltextsuche mit Eingabefeld und ggfs. Erweiterungen. Es können Texte und Bilder durchsucht werden. Für eine detailliertere Suche kann er Operatoren einsetzen, die bspw. die Suchziele definieren oder das Ranking konfigurieren. Jeder soll nach Definition von YaCy nachvollziehen können, wie Informationen für die Suchmaschine gewonnen und dargestellt werden.<sup>4</sup> (siehe Abbildung)

---

<sup>3</sup> <http://yacy.net/de/>

<sup>4</sup> <http://de.wikipedia.org/wiki/YaCy>

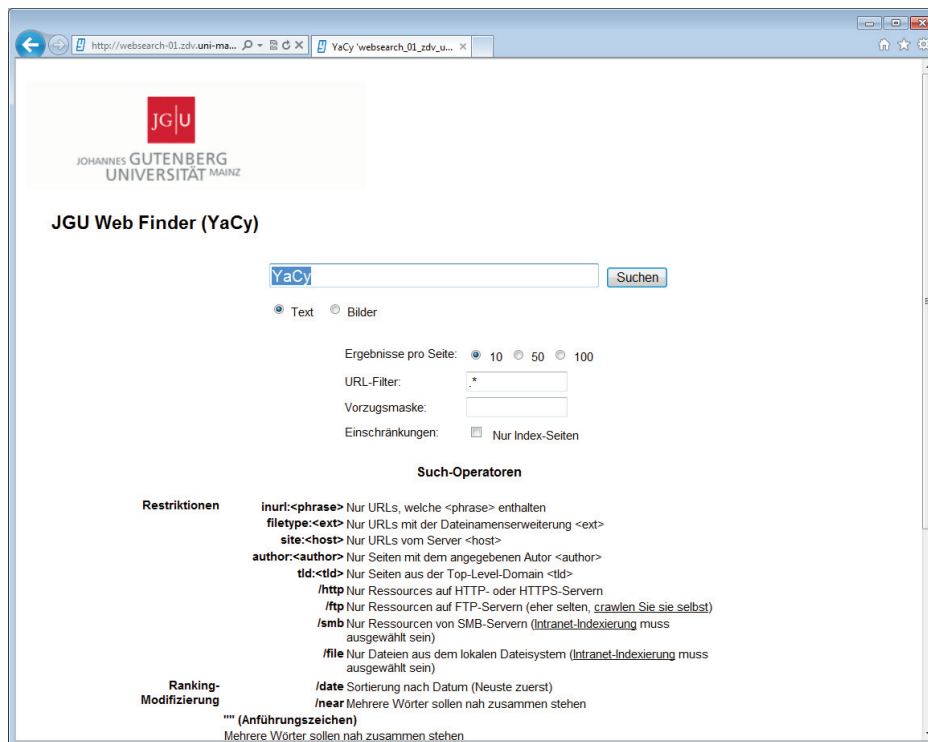


Abb.2. Such- und Darstellungsoptionen können definiert werden

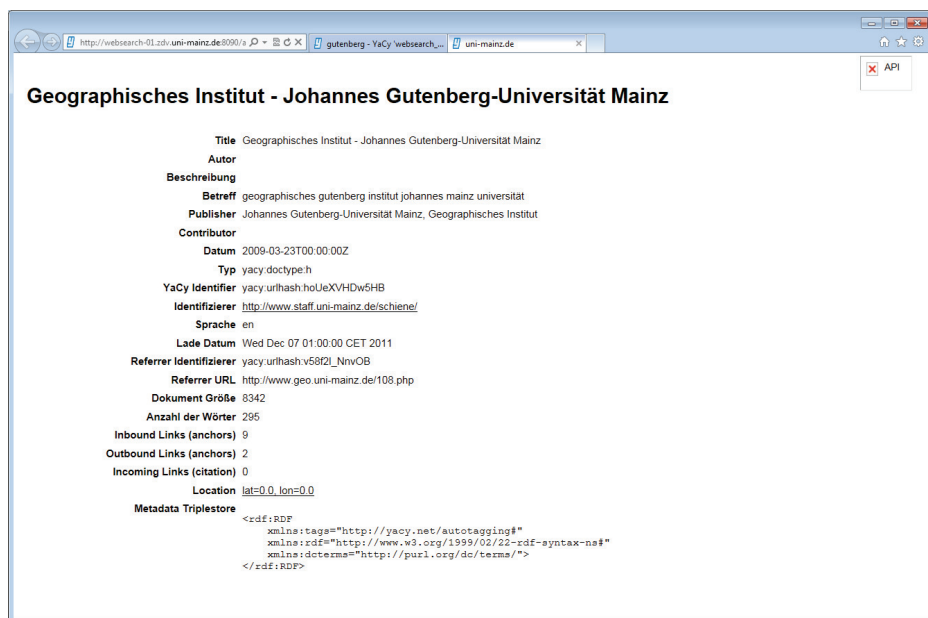


Abb.3. Metadaten eines Eintrags

Welche Vorteile haben Open Source Systeme wie YaCy gegenüber den kommerziellen Angeboten, welche Nachteile haben offene Systeme? YaCy und andere offene Systeme werden als Software in der Regel kostenlos angeboten. OpenSource Software kann darüber hinaus bearbeitet, weiterentwickelt und weitergegeben werden.<sup>5</sup> Mit den entsprechenden Kenntnissen kann die Software also genutzt werden, ohne dass spezielle Lizenzkosten entstehen. Völlig kostenlos ist der Einsatz von Open Source Software natürlich nicht. Die Hardware bereitgestellt, Implementierungs- sowie Entwicklungsressourcen müssen finanziert werden. Letzteres ist allerdings auch bei kommerzieller Software der Fall. Darüber hinaus bietet die offene, kollaborative Struktur von Software wie YaCy eine bessere Einsicht in die zugrundeliegenden Prozesse und Methoden, ein Punkt, der insbesondere bei kommerzieller Software oft kritisiert wird. Kommerzielle Systeme sind in ihren Lizenzen teuer bzw. im Leistungsumfang begrenzt.<sup>6</sup> Für kommerzielle Software steht allerdings durch Hersteller und Dienstleister insbesondere bei hoher Marktdurchdringung professionelle Unterstützung bereit. Die Unterstützung durch Dritte kann aber auch für quelloffene Software gefunden werden, hierfür entstehen dann aber ebenfalls Kosten.<sup>7</sup>

Welche Arten von Wissensquellen können durch spezifische Suchmaschinen erfasst werden, wie muss die qualitative Bewertung dieser Quellen erfolgen, die sich evtl. den herkömmlichen Kategorien der wissenschaftlichen Quellen entziehen? Spezifische Suchmaschinen bilden einen Filter für das Wissen, das der Zielgruppe zur Verfügung gestellt wird. Dieser Filter bezieht sich in der Regel allerdings nicht auf das einzelne Dokument sondern auf die zu indexierende Dokumentenquelle. Indexiert werden nur die Repositorien, die durch den Bibliothekar definiert werden. Die Auswahl kann dabei in Zusammenarbeit mit den Fachreferenten der Bibliothek, den Wissenschaftlern in den Fachbereichen und den Studenten erfolgen. Dabei können moderne Formen der Kommunikation wie offene Blogs und Diskussionsformen Anwendung finden.

---

<sup>5</sup> [http://en.wikipedia.org/wiki/Open-source\\_software](http://en.wikipedia.org/wiki/Open-source_software)

<sup>6</sup> [yacy.net/.../YaCy\\_CLT2011\\_Flyer\\_216x303.pdf](http://yacy.net/.../YaCy_CLT2011_Flyer_216x303.pdf)

<sup>7</sup> Kneifel, F. (2009): Mit Web 2.0 zum Online-Katalog der nächsten Generation. Seite 137. Dinges & Frick, Wiesbaden.



Hier von einem Paradigmenwechsel im Bild des Bibliothekars zu sprechen, ist Ansichtssache.<sup>8</sup> Zumindest werden tradierte Formen der Quellenauswahl und Kommunikation auf moderne Systeme umgestellt.

Bei der Auswahl der Quellen gilt es, die geltenden Nutzungsrechte zu beachten, nicht alles was indexiert werden kann, darf auch indexiert werden. Darüber hinaus gibt es wichtige Informationsquellen, die von der Suchmaschine nicht indexiert werden können. Denkt man an Bücher und Zeitschriften, können hier aber traditionelle, elektronische Kataloge genutzt werden, um indirekt den Zugang zu den Informationen zu ermöglichen.

#### Literaturverzeichnis

YaCy - Pressegespräch: Dezentrale Suche in Unternehmensnetzen – Yavid gegen Googliat ?,  
<http://www.dgi-info.de/index.php/aktuelles/456-yacy-pressegesprach-dezentrale-suche-in-unternehmensnetzen-yavid-gegen-googliat-> (27.2.2012)

Franziska Voß: ViFa-Websuche – Der neue Websuchraum der ViFa medien buehne film,  
<http://blogs.sub.uni-hamburg.de/webis/2012/05/10/vifa-websuche-der-neue-websuchraum-der-vifa-medien-buehne-film/> (10.5.2012).

Konrad Lischka: Suchmaschine YaCy - Deutsche Entwickler basteln dezentrales Anti-Google  
<http://www.spiegel.de/netzwelt/web/0,1518,800925,00.html> (1.12.2011)

DIY search engine takes on Google - Google, Bing, Yahoo and other search sites have a new rival called YaCy. <http://www.bbc.co.uk/news/technology-15935550> (29.11.2011)

Scott Goodson: A New Movement Against Google, Bing and Yahoo Search.  
<http://www.forbes.com/sites/marketshare/2011/11/30/a-new-movement-against-google-bing-and-yahoo-search/> (30.11.2011)

Simon Phipps: Amazon, P2P and non-centralised infrastructure,  
<http://blogs.computerworlduk.com/simon-says/2010/12/non-centralised-infrastructure/index.htm> (10.12.2010)

<http://de.wikipedia.org/wiki/YaCy>

<http://yacy.net/de/>

---

<sup>8</sup> Stühl-Strohmenger, W. (2008): Digitale Welt und Wissenschaftliche Bibliothek – Informationspraxis im Wandel. Harrassowitz Verlag, Wiesbaden.

## **Vernetzung mit MyCoRe – Eine Repository-Software vernetzt Systeme, Daten und Menschen**

Wiebke Oeltjen

Universität Hamburg

### **Zusammenfassung**

Die Vernetzung von Daten und Systemen wurde erst durch das Internet derart umfassend möglich, wie es heute Realität ist. Software, die online genutzt wird, ist dabei das verbindende Element. Am Beispiel der Repository-Software MyCoRe<sup>1</sup> sollen verschiedene Aspekte der Vernetzung von Menschen, Daten und Systemen vorgestellt werden. MyCoRe ist eine Open-Source-Software zur Entwicklung von Dokumenten- und Publikationsservern, digitalen Archiven oder institutionellen Repositorien. Der Name „MyCoRe“ weist darauf hin, dass es sich um einen Softwarekern (*My Core*) handelt, der in individuellen Anwendungen zum Einsatz kommt. Dieser Softwarekern wird bundesweit weiterentwickelt und an verschiedenen Standorten eingesetzt. Neue Entwicklungen von MyCoRe-Anwendungen sind Beispiele für die Verknüpfung von Bibliotheken. Doch nicht nur auf der Systemebene gibt es Vernetzung mit MyCoRe, sondern auch auf der Datenebene und in der MyCoRe-Community.

### **Abstract**

The tight and voluminous integration of data and systems we experience nowadays is directly coupled with the growth of the internet. Software that is used online is the connecting medium. Using the example of MyCoRe<sup>1</sup>, an open source repository software, a variety of issues concerning the network of people, data and systems will be introduced. MyCoRe is an open source software framework for the development of document and publication servers, digital archives and institutional repositories. The name “MyCoRe” refers to the circumstance, that the software is based on a core engine (*My Core*), which is used in custom applications. This software core is developed and used nationwide at various locations. Recent developments of MyCoRe applications are examples of the linking of libraries at technical, institutional and social level. It is important to notice that the technical links are not only regarded on the systemic but also on the data level of the MyCoRe applications and even further on a social level within the MyCoRe community.

---

<sup>1</sup> Link zur MyCoRe-Homepage: <http://www.mycore.de>

## Einleitung

In diesem Beitrag soll es um die Vernetzung von Daten, Menschen und Systemen am Beispiel der Open-Source-Software MyCoRe gehen. Der Kontext ist einerseits das Internet mit dem WorldWideWeb (WWW), in dem Web-Applikationen auf der Basis von MyCoRe benutzt werden können. Andererseits ist der Bezug in der realen Welt durch Bibliotheken mit Online-Diensten gegeben, die Daten erfassen, digitalisieren, katalogisieren, im WWW zugänglich machen und vernetzen. Doch zuvor soll die Software MyCoRe kurz vorgestellt werden. Das Akronym „MyCoRe“ bedeutet „*My Content Repository*“. Es weist darauf hin, dass man mit der Software ein eigenes Repositorium entwickeln kann. Dies können Dokumenten- und Publikationsserver oder institutionelle Repositorien sein. Auch können Archive, Kataloge oder sonstige Verzeichnisse mit digitalisiertem Inhalt auf der Basis von MyCoRe entwickelt werden (siehe auch Dobratz 2009). Der Inhalt der Repositorien kann sich stark unterscheiden: gespeichert werden Metadaten jeglicher Art, Kategorien, Klassifikationen, Verweise sowie Dokumente in unterschiedlichen Sprachen und Schriften, Bilder, Digitalisate, Audio- und Video-Dateien oder Dateiverzeichnisse, die auch hierarchisch sein können.

Ursprünglich bedeutete MyCoRe „*MILESS Community Content Repositories*“, was als Referenz an die Software „Miless“ zu verstehen ist, aus der heraus MyCoRe entwickelt wurde. Miles ist ein Dokumentenserver, der an der Universitätsbibliothek Duisburg-Essen betrieben und unter dem Namen „DuePublico“ kontinuierlich weiterentwickelt wird (vgl. Lützenkirchen 2002). Die Aussprache des Namens „*my core*“ weist auf eine weitere Bedeutung hin, dass es sich nämlich um einen Softwarekern handelt, der eine Klassenbibliothek bereit stellt. Sie ist in der Programmiersprache Java implementiert. Dieser Softwarekern stellt alle wesentlichen Funktionen bereit, die in Repositorien, Archiven oder Dokumentenservern zum Einsatz kommen. Die Möglichkeiten des Einsatzes und die Verwendung der Funktionen werden exemplarisch in der Beispielanwendung „DocPortal“<sup>2</sup> gezeigt. Weitere Details zum Konzept von MyCoRe sind die Anpassbarkeit des Metadatenmodells, der Klassifikationen und der Weboberflächen sowie die Konfigurierbarkeit der Such- und Eingabemasken.

---

<sup>2</sup> Die Webanwendung „DocPortal“ hat die Adresse <http://www.mycore.de:8291>,

Die Konfigurationen werden mit den Auszeichnungssprachen XML (*eXtensible Markup Language*) und XSL (*eXtensible Stylesheet Language*) vorgenommen. Der Quelltext der Software ist *Open Source*, das heißt, der Code ist unter der Lizenz „GPL“ (*GNU General Public License*<sup>3</sup>) frei verfügbar und steht auf der Softwareplattform *SourceForge*<sup>4</sup> zum Herunterladen bereit.

### **Vernetzung von Systemen: Repositorien auf der Basis von MyCoRe**

MyCoRe-Anwendungen sind an mehr als 20 Standorten in Deutschland installiert. Davon sind einzelne MyCoRe-Instanzen mit weiteren Bibliotheksbeständen vernetzt. Als Beispiel soll hier das Papyrus-Portal genannt werden, das in Kooperation mit Partnern am Rechenzentrum der Universität Leipzig entwickelt wurde (vgl. Freitag et. al 2012). Das Papyrus-Portal bietet eine einheitliche Rechercheplattform über deutschlandweit verteilte digitale Papyrussammlungen<sup>5</sup>. Beteiligt sind zehn digitale Bibliotheken in Bonn, Erlangen, Gießen, Halle, Heidelberg, Jena, Köln, Leipzig, Trier und Würzburg (siehe Abb. 1 mit der Startseite des Papyrus Portals).



**Abbildung 1: Das "Papyrus Portal"**

Was diese digitalen Bibliotheksbestände verbindet, das ist ein einheitliches Datenmodell und ein Standard für die Erschließungskategorien durch Verwendung gleicher Klassifikationen (vgl. Oeltjen 2010, S. 262). Mit Hilfe von Webservices ermöglicht die MyCoRe-Software eine gleichzeitige Suche über verteilt vorliegende Datenbestände (*federated search*). Bei einer Recherche kann selektiert werden, in welchen Bestän-

---

<sup>3</sup> Siehe „GNU General Public License (GPL)“ unter <http://www.gnu.org/copyleft/gpl.html>,

<sup>4</sup> Siehe MyCoRe auf der Projektseite der SourceForge-Plattform <http://mycore.sourceforge.net>

<sup>5</sup> Link zum Papyrus-Portal: <http://www.papyrusportal.de>. Das Projekt wurde von der Deutschen Forschungsgemeinschaft (DFG) finanziell unterstützt.

den gesucht werden soll. Die Suchergebnisse werden als eine Ergebnisliste angezeigt.

Detailansichten ermöglichen Forschern die Entzifferung, Auswertung und Deutung der in den Datenbanken gespeicherten digitalisierten Papyri, Ostraka oder anderen Dokumenten. Eine Ansicht der Digitalisate ist mit dem integrierten Bildbetrachter möglich (siehe Abb. 2).



**Abbildung 2: Bildbetrachter mit Darstellung eines Papyrus**

In der MyCoRe-Anwendung des Papyrus- und Ostraka-Projektes in Halle, Jena und Leipzig<sup>6</sup> ist es außerdem möglich, ausgewählte Ostraka in 3-D-Darstellung zu betrachten. Die im Browser angezeigten Objekte können mit einem Zeigegerät – z. B. mit der Maus – interaktiv gedreht werden, so dass die antiken Tonscherben von allen Seiten sichtbar sind (siehe Abb. 3). Dies ermöglicht eine weitergehende Untersuchung und Auswertung des digital erfassten Forschungsmaterials. In dem Projekt wird explizit darauf hingewiesen, dass in den Metadaten Links zu weiteren Datenbanken und Katalogen erfasst sind, die ergänzende Informationen zu den Daten liefern. So sind die Daten in den MyCoRe-Anwendungen mit Papyrus- und Ostrakon-Abbildungen weitläufig vernetzt.

---

<sup>6</sup> Siehe <http://papyri-leipzig.dl.uni-leipzig.de>. Das Projekt wurde von der Deutschen Forschungsgemeinschaft (DFG) finanziert.



Abbildung 3: Ostrakon im 3-D-Betrachter des Papyrus- und Ostraka-Projektes Halle, Jena und Leipzig

### Vernetzung von Daten – Hyperlinks und Identifier

Das Hauptmerkmal des WorldWideWeb (WWW) ist die Vernetzung von *Webseiten* mittels Querverweisen (*Hyperlink*). Diese unidirektionalen Verweise werden als Sprungadressen auf dieselbe oder eine andere Webseite oder Internetressource selbstverständlich zur Vernetzung eingesetzt. So werden auch in institutionellen Repositorien oder Online-Archiven Hyperlinks als „*Uniform Resource Locator*“ (URL) in den Metadaten erfasst. Sie verweisen zum Beispiel im Kontext bibliografischer Daten auf die Internetpräsenz eines Verlages oder auf weiterführende Online-Ressourcen. Die Verlässlichkeit in der Erreichbarkeit des verlinkten Objekts ist dabei ebenso wichtig, wie die korrekte Angabe der Adresse. Auch muss die Dauerhaftigkeit der Identifier sicher gestellt werden, was mittels „*Persistent Uniform Resource Locator*“ (PURL), „*Digital Object Identifier*“ (DOI) oder „*Uniform Resource Name*“ (URN) sichergestellt werden kann. In MyCoRe-Anwendungen werden solche persistenten Verweise verwendet, um Metadaten und Daten zu vernetzen.

Eine weitere Art der Vernetzung im WWW gewinnt zunehmend an Bedeutung. Das ist die Vernetzung von Daten als „*Linked Data*“ und „*Linked Open Data*“. Bei dieser Art Vernetzung werden an Hand von Identifier Links generiert und Daten automatisiert in Webseiten eingebunden und dargestellt. Als Beispiel sei hier die MyCoRe-Anwendung „Rostocker Professorenkatalog“<sup>7</sup> genannt, die über das BEACON-Format Personendaten mit digitalen Ressourcen verknüpft. Der Schlüssel

<sup>7</sup> Siehe dazu den Beitrag in diesem Band von R. Stephan, K. Labahn, M. Glasow: „Vernetzung biographischer Online-Ressourcen – Catalogus Professorum Rostochiensium und Rostocker Matrikelportal“. Der Link zum Rostocker Professorenkatalog „Catalogus Professorum Rostochiensium“ lautet: <http://cpr.uni-rostock.de>

ist ein eindeutiger Identifikator, wie er z.B. mit der Nummer in der gemeinsamen Normdatei (GND) gegeben ist. Die GND vereint u.a. die Normdateien Gemeinsame Körperschaftsdatei (GKD), Personennamendatei (PND) und Schlagwortnormdatei (SWD) in einer gemeinsamen Normdatei<sup>8</sup>. Personendaten, die bei der Deutschen Nationalbibliothek erfasst und mit einer GND-Nummer versehen sind, können nun automatisiert in den Professorenkatalog eingefügt werden.

Die Vernetzung der Daten verlangt, dass Standards eingehalten werden, damit die Austauschbarkeit gewährleistet ist. Deshalb ist es für den Datenaustausch von Bedeutung, standardisierte Datenmodelle zu verwenden, wie z.B. XMetaDissPlus oder Epicur. Dazu gehört auch, dass standardisierte Klassifikationen wie z.B. die Dewey-Dezimalklassifikation (DDC) verwendet werden. In der MyCoRe-Anwendung DuePublico<sup>9</sup> werden wie in anderen MyCoRe-Anwendungen auch die Formate implementiert. Diese elektronischen Dokumente werden über das OAI-Protokoll, dem „*Open Archives Initiative Protocol for Metadata Harvesting*“<sup>10</sup> (OAI-PMH 2.0) online ausgeliefert. Die OAI-Schnittstelle ist in MyCoRe-Anwendungen generisch und erweiterbar implementiert, so dass Repositorien auf der Basis von MyCoRe als OAI-Data-Provider dienen und Metadaten zum Harvesten ausliefern können. In der Beispielanwendung DocPortal ist die OAI-Schnittstelle integriert und kann online getestet werden<sup>11</sup>.

Das Metadatenformat „Metadata Object Description Schema“ (MODS<sup>12</sup>) ist lange bekannt: Es wurde vor zehn Jahren als XML-Schema von der „Library of Congress“ (LoC) entwickelt und ist gut dokumentiert. Entwickelt wurde es für XML-Anwendungen; es eignet sich somit für Langzeitlösungen („*long-term solutions*“) im Bibliotheksumfeld (vergl. McCallum 2004, S. 83). Es dient zur Beschreibung von bibliografischen Metadaten und wird zum Beispiel zur Beschreibung historischer Bestände eingesetzt. Nun wurden in der MyCoRe-Community erste Anwendungen auf der Basis eines MyCoRe-Kerns implementiert, die das MODS-Metadatenformat im Datenmodell verwenden. Die Funktionalität umfasst sowohl den Import von MODS-Dokumenten (in der MODS-Version 3.4) als auch die Ausgabe im MODS-Format.

---

<sup>8</sup> Siehe Deutsche Nationalbibliothek: „Gemeinsame Normdatei (GND)“.

[http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd\\_node.html](http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html)

<sup>9</sup> Der Link zum Dokumentenserver DuEPublico: <http://duepublico.uni-duisburg-essen.de>

<sup>10</sup> Link zum OAI-PMH: <http://www.openarchives.org/pmh/>.

<sup>11</sup> Der Link <http://www.mycore.de:8291/oai2?verb=Identify> liefert ein OAI 2.0-Anfrageergebnis.

<sup>12</sup> Link zur Library of Congress: <http://www.loc.gov/standards/mods/>

Da die Metadaten in MyCoRe in dem universellen Datenaustauschformat XML gespeichert werden, ist es leicht möglich sie in andere Datenformate zu transformieren und dann zu exportieren. Datensätze, die als Ergebnisliste einer Recherche vorliegen, können in verschiedene Formate ausgegeben werden: Sie können direkt als HTML-Seite angesehen werden, als PDF-Dokument gespeichert oder in bibliographischen Formaten wie z.B. BiBTeX oder EndNote weiterverarbeitet werden.

Für die Vernetzung von Daten ist nicht nur deren Export von Bedeutung, sondern auch der Import. Neben einer einfachen Schnittstelle für den Batch-Import von XML-Dateien ist z.B. eine Vernetzung mit Systemen möglich, die über die SWORD-Schnittstelle Metadaten und Dokumente ausliefern. Das Akronym SWORD steht für „*Simple Web-service Offering Repository Deposit*“<sup>13</sup>. MyCoRe-Anwendungen können als SWORD-Client konfiguriert werden, so dass sie Metadaten und Dateien auch mit komplexen Datenstrukturen (z.B. als zip-Dateien) automatisiert importieren können. Genutzt wird das automatisierte Ausliefern von Metadaten und Dokumenten zur Zeit von der MyCoRe-Anwendung DuePublico, die u.a. Daten von „*Biomed Central*“ erhält<sup>14</sup>.

Ein weiterer Aspekt der Vernetzung ist es, Suchmaschinen den Zugang zum Datenmaterial zu ermöglichen und sie über den Inhalt des Repositoriums oder Archivs zu informieren. Damit die Metadaten und Dokumente der Repositorien von Suchmaschinen gefunden und erfasst werden können, ist es optional möglich das Sitemaps-Protokoll in MyCoRe-Anwendungen zu verwenden. Eine weitere Möglichkeit, Suchergebnisse zu verbessern, ist das Generieren von Indexseiten, was ebenfalls mit MyCoRe-Anwendungen möglich ist.

### **Vernetzung von Menschen: Die MyCoRe-Community**

Die kontinuierliche Weiterentwicklung der Open-Source-Software MyCoRe wäre nicht ohne Vernetzung möglich. In der MyCoRe-Community, die aus MyCoRe-Entwicklern an Universitätsbibliotheken und -Rechenzentren sowie aus MyCoRe-Nutzern besteht, wird kooperativ und bundesweit verteilt an dem Softwarekern gearbeitet. Der Austausch wird durch regelmäßige Treffen, Telefonkonferenzen, E-Mail-Austausch und Nutzung von Chat-Räumen gepflegt. Koordiniert wird der

---

<sup>13</sup> Siehe <http://swordapp.org/> für Informationen zu SWORD.

<sup>14</sup> Weitere Informationen dazu auf der Seite von „BioMed Central“: <http://www.biomedcentral.com/libraries/aad>



Kontakt über die MyCoRe-Geschäftsstelle, die am Rechenzentrum der Universität Hamburg angesiedelt ist.

Die Entwickler des MyCoRe-Softwarekerns greifen auf eine Software-Infrastruktur zurück, die ihnen Werkzeuge und Softwarekomponenten liefert. Da sind nicht nur die Programmiersprachen zu nennen, sondern auch die vielen Werkzeuge, die qualitativ gute Softwareentwicklung unterstützen. Im MyCoRe-Entwicklungsprozess ist dies z.B. das Softwareintegrationswerkzeug Jenkins. Das ist ein webbasiertes System zur kontinuierlichen Integration von Softwarekomponenten. Die MyCoRe-Entwickler handeln nach den Grundsätzen der Open-Source-Community, die erlaubt, dass die Software beliebig kopiert, verbreitet und genutzt wird und die verlangt, dass die Software verändert und in der veränderten Form weitergegeben werden darf.

### **Ausblick**

Vernetzung stellt für digitale Bibliotheken einen Mehrwert dar. Die einmal implementierten Systeme werden durch Anbindung über ihre Schnittstellen zu einem Netzwerk, in dem Daten ausgetauscht und verknüpft werden. Die Nutzerinnen und Nutzer profitieren von der Vernetzung durch eine größere Verfügbarkeit von Daten und gleichzeitig sparen sie durch Vernetzung, weil redundante Codeerzeugung vermieden und doppelte Datenerfassung vermindert wird. Die MyCoRe-Community hat sich ein soziales Netzwerk geschaffen, in dem die MyCoRe-Software kontinuierlich weiterentwickelt wird. Sie nutzt eine Software-Infrastruktur, die hilft, mit knappen Ressourcen dennoch viel zu bewirken. Ziel für die Weiterentwicklung ist unter anderem die Integration von Solr als Lucene-basierte Suche. Ein weiterer Baustein ist die Neugestaltung des DocPortal-Layouts. In der MyCoRe-Beispielanwendung DocPortal soll das CSS-Framework von Twitter Bootstrap integriert werden. An verschiedenen Standorten entstehen neue MyCoRe-Anwendungen und bestehende werden weiter entwickelt, vernetzt oder nachgenutzt, wie dies z.B. für den Rostocker Professorenkatalog der Fall ist. Ziel ist es darüber hinaus, je nach Nachfrage und Bedarf auch neue Datenmodelle zu implementieren und Schnittstellen für den Austausch der Daten zu integrieren. Nicht zuletzt ist es ein Ziel, weitere MyCoRe-Anwender zu gewinnen.

### Quellen der Abbildungen

Abbildung 1: Das "Papyrus Portal". <http://www.papyrusportal.de>

Abbildung 2: Bildbetrachter mit Darstellung eines Papyrus.  
[http://papyri-leipzig.dl.uni-leipzig.de/receive/HalPapyri\\_schrift\\_00002440](http://papyri-leipzig.dl.uni-leipzig.de/receive/HalPapyri_schrift_00002440)

Abbildung 3: Ostrakon im 3-D-Betrachter des Papyrus- und Ostraka-Projektes Halle, Jena und Leipzig. Digitales Dokument : OLipsInv0747.html, siehe Metadaten: [http://papyri.uni-leipzig.de/receive/UBLDocument\\_document\\_00005980](http://papyri.uni-leipzig.de/receive/UBLDocument_document_00005980)

### Literaturverzeichnis

Dobratz, Susanne: Open-Source-Software zur Realisierung von Institutionellen Repositories – Überblick. Erschienen in: ZfBB 54 (4-5) 2007, S. 199-206  
<http://nbn-resolving.de/urn:nbn:de:kobv:11-10081380>

Freitag, Stefan; Gerhardt, Marius; Kupferschmidt, Jens; Scholl, Reinhold : Papyrus Portal. Umsetzung einer Portallösung für alle Papyrussammlungen in Deutschland. Version 1.4, Leipzig, 15.03.2012, [http://www.papyrusportal.de/PapyrusPortal\\_Dokumentation.pdf](http://www.papyrusportal.de/PapyrusPortal_Dokumentation.pdf)

Lützenkirchen, Frank (2002): MyCoRe - Ein Open-Source-System zum Aufbau digitaler Bibliotheken. Datenbank Spektrum, 2(4), November 2002, 23-27, [http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-11198/MyCoRe\\_DB-Spektrum.pdf](http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-11198/MyCoRe_DB-Spektrum.pdf)

McCallum, Sally H. (2004): An introduction to the Metadata Object Description Schema (MODS). Library hi tech 22(1), 82–88, doi:10.1108/07378830410524521

Oeltjen, Wiebke (2010): Virtuelle Bibliotheken flexibel gestalten. In: Mittermaier, Bernhard (Hg.): eLibrary - den Wandel gestalten. Proceedings of the WissKom 2010, Schriften des Forschungszentrums Jülich / Reihe Bibliothek/Library, Vol. 20, Zentralbibliothek Verlag, S. 259 – 266

Alle in diesem Beitrag genannten Web-Adressen und Quellen wurden am 6.9.2012 abgerufen und geprüft.

---

## **10 years collaborative catalogue enrichment - the dandelon.com community**

Manfred Hauer

AGI – Information Management Consultants, Neustadt / Weinstrasse

### **Zusammenfassung**

Seit 2002 haben Autoren in OPACs endlich etwas mehr zu sagen - durch Digitalisierung von vorwiegend Inhaltsverzeichnissen und deren computerlinguistischer Extraktion ergänzen deren Begriffe die Welt der kontrollierten Terminologien bei der Suche im Katalog. 2003 entstanden daraus das offene Netzwerk von Bibliotheken und deren kollaborativer Online-Dienst "dandelon.com" zum Austausch zwischen den Bibliotheken und zur Suche für die Öffentlichkeit. Schließlich liefern Dokumentationszentren ihre Thesauri zur Query-Expansion und zur Navigation durch die Fachterminologien. Viel Software, viele Bibliotheken, zahlreiche Verlage und einige Dokumentationszentren sind darin vernetzt: Daten, Menschen und Systeme. Catalogue Enrichment ist "Brückenlösung", "Migrationspfad" zu neuen, volldigitalen, semantisch vernetzten Sammlungen - was dem Naturwissenschaftler seine Messdaten, sind dem Gesellschaftswissenschaftler nicht zuletzt Zeitungen, Protokolle mit Bildern, Namen, Ereignisse, Zusammenhängen - wie sie im Umfeld von dandelon.com zunehmend entstehen und mittelfristig soweit möglich dort zusammenwachsen werden.

### **Abstract**

Since 2002 authors got a stronger voice in library management systems by adding searchable tables of contents and extracting with text mining technologies major keywords and phrases. 2003 a search and exchange platform names "dandelon.com" was added. Data from libraries, publishers and terminologies from documentation specialists merged into one solution for multilingual, semantic retrieval. Multiple solutions with different contents, formats and documents arose from it.

*Eine kleine, wahre Geschichte – Bibliothekartag 2008 in Mannheim: Frau B. hatte gerade ihr Diplom in Geologie an der Universität Mannheim beendet als wir sie für die Tagung engagierten. Am Abend vor ihrem ersten Arbeitstag bat ich sie, das Thema ihrer Abschlussarbeit in dandelon.com zu recherchieren. Völlig überrascht berichtete sie am nächsten Morgen, dass sie zuvor fünf Monate lang nach den sehr wenigen einschlägigen Aufsätzen gesucht hatte. In dandelon.com hatte sie sofort die richtigen Konferenzbände gefunden. Es ging um „Die letzte Eiszeit in Südafrika“. 5 Monate oder 5 Minuten – darüber sprechen wir.*

## **1. Unsere Fiktion von der gleichen Welt**

An manchen sonnigen Tagen ist die eine Hälfte des Mondes und die zugleich die Sonne mit bloßem Auge wunderbar zu sehen und die zweite Hälfte der Kugel zu erahnen. Bis zu dieser Ahnung benötigte die Menschheit über 100.000 Jahre. Unsere egozentrische Sicht auf die Welt und die Fiktion, andere sähen sie genauso, ist täglich Grund für Dialog, Verwunderung, Missverständnis, Auseinandersetzung bis hin zu kriegerischem Mord und Totschlag. Niemand kann die Welt durch meine Augen sehen, denn das Gesehene entsteht wesentlich erst durch die Interpretation des Gehirns mit stets individueller Modellierung. Über Sprachen, Abbildungen und Aktionen bemühen sich Tiere und Menschen um einen Konsens über das sinnlich Erfahrene. Auch Bibliothekare sind von dieser Fiktion keineswegs frei.

Gerne lasse ich in meinen Hochschulkursen zum Thema Indexierung gleiche Medienobjekte (Zeitungsartikel, also eher kurze Texte) verbal beschreiben. Die Reihenfolgen, ein Indikator für Relevanzbeurteilung, klaffen weit auseinander und die Worte selbst (Schreibvarianten und Synonyme werden nicht als solche gewertet) stimmen rund 25 % im Schnitt überein. Dort wo der Sprachraum künstlich stark verknüpft, Synonyme zusammengeführt sind auf flache Listen, hierarchische Klassifikationen, polyhierarchische Thesauri oder Ontologien, die meist sehr mühsam ausgearbeitet wurden, steigt das Maß der Übereinstimmung zwischen wenigen kompetenten Sprechern – nicht aber Bibliothekskunden -, kommt aber bei einer dem Gegenstand angemessenen hinreichenden Differenzierung kaum in Richtung 70 %. Maschinelle Verfahren, die von Menschen lernen, sind auch nicht besser als ihre Lehrer. Dummerweise wollen sich die Benutzer der Datenbanken diesen verbalen Verknüpfungen nicht aktiv anschließen. Google & Co haben ihnen beigebracht, dass das nicht sein muss. Der Bibliothekar/Dokumentar schiebt sich mit seiner Beschreibung zwischen den Autor und den nächsten Autor, denn die Mehrzahl der Bibliotheks- und Datenbankbenutzer lesen überwiegend fachliche Literatur zum Zwecke einer eigenen Publikation ihres Wissens. Die fachlich spezialisierten Autoren untereinander verstehen sich meist besser, als den längst aus der aktiven Forschung- und Entwicklung ausgestiegenen Bibliothekar mit seiner verknüpften Sprache. Diese verknüpfte Sprache kommt systematisch mit den

Wortschöpfungen und Uminterpretationen von Begriffen zeitlich nicht mit, da sie sich gerade um Standardisierung bemühen muss.

Weil viele Autoren nicht mehr besitzen als dieses mühsam erarbeitete Wissen und dieses für andere oft sehr nützlich ist, gelang es ihnen, dieses graduell zu schützen, um damit notwendige wirtschaftliche Einkünfte zu erzielen. Die Verlage wurden zu ihren Agenten. Dieser wirtschaftliche sinnvolle Schutz behindert andere Wirtschaftssubjekte in der Auswertung. Open Access, Linked Data sind junge Ansätze von überwiegend staatlich bezahlten Akteuren, die Verwertungslogik entsprechend den großen Fortschritten in einer globalen IT-Vernetzung neu zu gestalten. Im Buchmarkt dominieren nach wie vor die in 500 Jahren entstandenen Verlage mit ihrem Geschäftsmodell und eher Mitspieler als Team Leader sind der Buchhandel und die klassischen Bibliotheken.

## **2. Nischen implementieren**

Erstens: Als Student der Soziologie und Politikwissenschaften in Konstanz liebte ich das stundenlange Browsen entlang der Regale, auch in „fremden“ Fachbereichen, um jeweils die Inhaltsverzeichnisse zu überfliegen. Zweitens: 1981 lernte ich, nun im Aufbaustudium Informationswissenschaft unter Rainer Kuhlen, Udo Hahn und Ulrich Reimer, dass maschinelle Verfahren zum Textverstehen grundsätzlich in der Lage sind, auch große Textkorpora deutlich schneller als Menschen zu analysieren. Drittens: genau die Kombination aus beidem ist heute mein Beruf. Wir analysieren – vorwiegend – Inhaltsverzeichnisse aus derzeit 53 Sprachen, um sie maschinell zu beschreiben.

Bis 2002 war die intellektuelle Beschreibung mit durchschnittlich 4-8 Fachbegriffen, teils aus dem Titel, teils aus der verknüpften Terminologie, State-of-the-Art im Bibliothekswesen. Die meisten Bibliotheken weltweit sind noch immer auf dem Stand vor 10 Jahren. Wohl diskutieren sie über neue Formate, kaum über Inhalte.

Erstaunlich, denn die Wissenschaftler, die sich seit mindestens 250 Jahren gegenseitig über Briefe Zusammenfassungen der gelesenen Aufsätze lieferten, denn schon damals empfanden sie einen „Information Overflow“, was schließlich in die Profession der Dokumentare sich entwickelte, nutzen längst das Konzept der

Zusammenfassung, das weit mehr als 8 Fachbegriffe in der Regel lieferte. Die bis heute so weit verbreiteten Limits sind die Spätfolgen der Karteikarten – heute eben in digitaler Form, aber inhaltlich nicht reicher. Information Retrieval-Systeme überwand in den 80er Jahren die lange zu hohen Kosten der Speichersysteme und öffneten die Suche im gesamten Dokument. IBM und andere Rechnerhersteller waren die Pioniere, Google lernte zusätzlich Garfields Logik des Science Citation Index, nutzte das Internet zur Datenakquisition und ist heute der größte Gewinner daraus.

Die Einsichten aus Konstanz führten mich zweimal nach Saarbrücken, zunächst 1996 zu Harald Zimmermann und IDX, daraus erwuchsen erste Implementierungen maschineller Indexierung im Bibliothekswesen zuerst bei der UB Düsseldorf und 1999 zu dem von Zimmermann mitbegründeten An-Institut IAI, wo sein früherer Mitarbeiter Dieter Maß und Institutsdirektor Hans Haller sowie Paul Schmidt unter dem Namen Autindex das Verfahren neu implementierten. Haller ist Experte für maschinelle Übersetzung. Heute hat die Text Mining-Welt stark zugelegt, probabilistische Verfahren ersetzen oder ergänzen die linguistischen Verfahren – Google Translator ist derzeit das prominenteste Resultat der Sprachtechnologien.

Inhaltsverzeichnisse können Autoren und Verlage als Referenzinformation auf den eigentlichen Text in der Regel nicht schützen. 2007 erklärt der Börsenverein den Bibliotheken auch den „Nicht-Angriffspakt“. 2001 ersinnen Manfred Hauer, Autor des Beitrags und Karl Rädler, bis dahin schon mehrfach Pionier im Bibliothekswesen, von der Vorarlberger Landesbibliothek die Entwicklung eines Verfahrens zur Digitalisierung von Inhaltsverzeichnissen und maschinellen Extraktion dieser Texte und Übertragung der Indexate in deren ALEPH-Bibliothekssystem. Im Januar **2002** startete die Produktion mit **intelligentCAPTURE**. Die Nische war geboren, die Sprache der Autoren, konzentriert auf die im Schnitt 100 bis 500 wichtigsten Fachbegriffe, war jetzt im OPAC recherchierbar. Leider, wie sich rasch herausstellte ohne das Relevanz-Ranking wie bei modernen Retrieval-Systemen üblich. Die Recherchen in dem Produktionssystem intelligentCAPTURE ergänzten nicht selten die Recherchen der Bibliothekare im Katalog, da wegen der verwendeten Basis-Technologie Lotus Notes & Domino von IBM hier bereits Relevanz-Ranking und zusätzlich die Suche im Volltext seit 1994 Standard waren.

Damit alle Zugang zu dieser Funktionalität haben können, stellte AGI auf dem IFLA-Kongress **2003** erstmals die Anfänge von <http://www.dandelon.com> vor, mit kaum mehr als 3000 Inhaltsverzeichnissen, welche die Vorarlberger Landesbibliothek, Liechtensteinische Landesbibliothek und Universitätsbibliothek St. Gallen lieferten. Das Ziel der Entwicklung war von Anfang an, nicht „Bilder“ anzuzeigen, sondern in Bibliotheken modernes Information Retrieval einzuführen, also zumindest den OPAC zu ergänzen, besser aber neue Suchsysteme einzusetzen. Inzwischen ist die Idee der neuen Retrieval-Systeme unter dem Label „Discovery Engine“ bei Bibliotheksverbünden und immer mehr Bibliotheken angekommen – doch ohne Texte der Autoren wie Inhaltsverzeichnisse, Zusammenfassungen, Besprechungen nur die halbe Lösung.

Dandelon.com ist neben seiner noch immer kostenlosen, öffentlich frei zugänglichen Suche heute die Austauschplattform für alle angeschlossenen Bibliotheken. Parallel dazu tauschen die Verbünde untereinander aus, die Kreise überschneiden sich. Die erste öffentliche Vorstellung auf der IFLA fand sehr bescheiden auf dem Stand von ImageWare statt, die später zum wichtigsten Mitbewerber sich zeitweise entwickelten. Die Einschätzung von mehreren Anbietern für Bibliotheken, das eine sehr lange Durststrecke zur Durchsetzung zu überwinden sei, erwies sich als wahr.

### **3. Capture, Process and Distribute**

Obwohl das inzwischen mobile „intelligentCAPTURE mobile“ vorwiegend über seine gescannten Inhaltsverzeichnisse bekannt ist, ging es immer darum, für die Indexierung maschinell verwertbaren Content zu erfassen (capture). Über das klassische Bibliothekswesen entstand 2004 in Bregenz ein Parlamentsinformationssysteme mit dem Input Word-Dateien und ähnlicher Suchtechnik wie dandelon.com, zwei Stunden nach Sitzungsschluss sind die alle Protokolle öffentlich, maschinell und gemeinsam von Landtag, Landesbibliothek und Landesarchiv ergänzt und in bis zu 16 Sprachen suchbar. Im März 2005 schließen AGI-IMC und GBV einen Kooperationsvertrag über die Nutzung der dandelon.com-Inhaltsverzeichnisse im Katalog der des GBV und allen lokalen Katalogen der Mitglieder und umgekehrt hosted der GBV die AGI-Server. Dadurch kommen wir 2006 in eine vom GBV finanzierte Produktion und Evaluation, dürfen an den



Ausschreibung teilnehmen, gewinnen die meisten und arbeiten in der Folge für die TIB Hannover u.a. Im zweiten dieser Projekte, an der SUB Hamburg, stellen wir intelligentCAPTURE erstmals auf mobile Beine zur Arbeit zwischen auch engen Regalen – also ganz ohne Aushebung auf Bücherwagen. Heute ist diese Art von Projektgeschäft ein wichtiger Anteil am Umsatz und der Produktion in Deutschland.

2007 wurde die Suche der Landeshomepage Vorarlberg darauf umgestellt mit einem neu entwickelten, bürgernahen e-Government-Thesaurus dahinter. Ab 2004 sammelt die Vorarlberger Landesbibliothek in dandelon.com die Current Content-Daten von Swets und öffnet damit den Bereich „Aufsätze in Sammelwerken“. Dies deckt nur 50 % aller Zeitschriften ab - und weil die Buchinhaltsverzeichnisse retrospektiv nun abgearbeitet waren, beginnt die VLB ab 2009 Current Content-Daten für den Rest selbst zu erarbeiten. Bisher sind es knapp 1 Million Aufsatztitel wo möglich mit Abstract. Auch hier ist das Ziel, alle bisher relevanten Aufsätze in der Bibliothek nachzuweisen, also auch retrospektiv. Die Dandelon.com-Artikel-Suche mit automatischer Query-Expansion auf Basis von vielen parallel abgefragten Thesauri verbirgt sich weiterhin hinter der in Bregenz neuen Discovery-Lösung Aquabrowser, die 2012 die ALEPH-Oberfläche ersetzte.

Seit 2009 produziert AGI für die Deutsche Nationalbibliothek Inhaltsverzeichnisse, seit 2012 exklusiv und implementierte u.a. einen OAI-Server innerhalb des Domino-Servers und die vielen Benutzern bekannten Links auf die DNB-Datensätze und deren Logo im PDF. 2010 startete die digitale Bibliothek der Liechtensteinischen Landesbibliothek mit sämtlichen bis Ende 2005 publizierten Zeitungen (über 500.000 Seiten), langen Buchreihen und kompletten Zeitschriften – alles, auch Fraktur, gut suchbar und erstmals mit Lucene als Search-Engine hinter dem Domino-Server. Während die Zeitungen preisgünstig seitenweise erschlossen werden, wird bei Büchern und Zeitschriften die artikelweise Erschließung gewählt – jeweils mit Highlighting der Suchbegriffe in der Online-Abfrage im PDF. Das PDF-Format kann nicht nur auf mehrere kostenlose Browser zurückgreifen, sondern da es nie darum ging, nur retrospektiv Zeitungen darzustellen, sondern stets möglichst tagesaktuell zu werden, entspricht dies den Formaten der allermeisten Zeitungsverlage und zunehmend den Dokumentationsstandards öffentlicher Verwaltungen (in der offengelegten, standardisierten Version PDF/A). Die Implementierung ist so gewählt, dass grundsätzlich später neue Formate und noch bessere Erschließungen aus den Rohdaten generiert werden können.

intelligentCAPTURE hat also Verfahren zur Erfassung von Papier über Digitalisierung, strukturierten und unstrukturierten Daten. Die implementierte Scanlösung und ab 2006 standardisierten Flachbett-Scannern von (meist) Fujitsu führten zu besseren OCR-Resultaten, die wesentlich Grundlage für maschinelle Indexierung und Volltext-Retrieval, als bei vielen anderen und erlaubt mit heutigen Geräten und motivierten Mitarbeitern einen Durchsatz von über 10.000 manuell erfassten Inhaltsverzeichnissen mit internem Titelblatt pro Monat bei bis zu 60 Bücher pro Stunde. Nebenaufgaben wie Prüfungen im Katalog, Druck- und Einkleben von Barcodes oder Chips und weiträumige Büchertransporte und Korrekturen in der Aufstellung können die Arbeit projektspezifisch ergänzen.

Für die OCR wird für bisher real ca. 120 Sprachen Abbyy Fine-Reader mit gutem Resultat verwendet. Die neue Architektur von intelligentCAPTURE 8, ab September 2012 verfügbar, macht den Aufwand zur Integration für zusätzliche spezielle OCR-Programme überschaubar. Aktuell sind wir mit den OCR-Resultaten für arabische Texte, einer der schwierigsten Nüsse für Schrifterkennung, noch nicht entschieden.

Für die multilinguale Indexierung, 53 Sprachen seit September 2012/Version 8 inklusive Deutsch, übersetzen wir 52 Sprachen nach Deutsch und verwenden IAI Autindex in unserem CAI-Modul, Modul für Computer Aided Indexing, weil es ohne Training funktioniert und seit 10 Jahren ordentliche Resultate liefert, welche einige, nicht alle unsere Bibliotheken ganz oder teilweise in ihre Kataloge übernehmen. Leider ist Autindex nicht offen für unsere eigene Weiterentwicklung, Änderungen stehen für Release 9 an.

Die fertigen Ergebnisse werden nach MARC, MAB, OAI und anderen Formaten an die jeweiligen Bibliothekssysteme oder zwischengeschaltete DMS-Systeme übertragen und im Falle von Inhaltsverzeichnissen meist auch zur öffentlichen Suche und zum Austausch zwischen den Bibliotheken in dandelon.com gesammelt. Die Zielsysteme und Formate sind für die jeweiligen Workflows konfigurierbar. Der GBV nutzt für einen Teil der Kollektion unsere Domino-Datenbanken als DMS.

#### **4. Marktanteile**

Die Idee, Inhaltsverzeichnisse in Katalogen anzuzeigen und möglichst auch suchbar zu machen, wurde mehrfach geboren. 1996 realisierten wir für die Buchhandelsgruppe Schweitzer Sortiment bereits eine funktionierende Lösung inklusive Volltextsuche für 2,4 Millionen bibliografische Daten, damals mehr als Amazon und etwas über 300 Inhaltsverzeichnissen im Schweitzer Vademecum. Schweitzer erzwang einen Aufsichtsscanner, der de facto so lausige Images produzierte, dass keine damalige OCR daraus gute Resultate generieren konnte, doch Haupthindernis für den Verkauf war die Entfernung der Plastikhüllen – der Nachweis für ein neues Buch. Deshalb endete das Projekt bezüglich Inhaltsverzeichnissen während der Geburt. Der internationale Buchhändler Casalini Libri hat diese Idee mit unserer Technik für italienische, französische und spanische Titel seit Jahren im Einsatz.

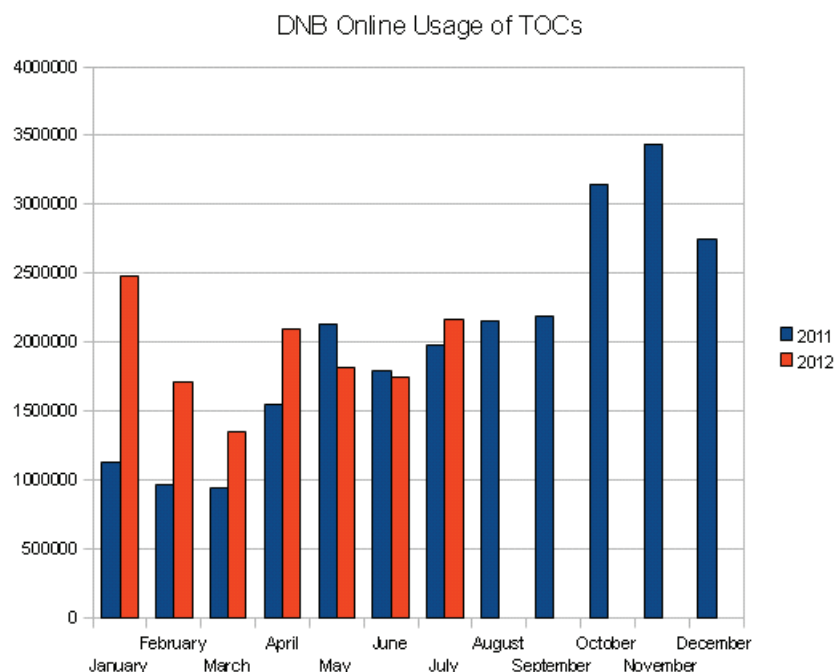
Zeitgleich zum Neustart der Idee 2001/2002 in Bregenz gab es zwei oder drei Projekt in den USA, die heute in die Aktivitäten der Library of Congress und von ProQuest / Bowker/ Syndetics/ Serial Solutions mündeten. Dort stand nicht das Scannen, sondern die Übernahme von Verlagsdaten im Vordergrund. Solches wäre hier durch uns mit tausenden von kleinsten Verlagen nicht durchsetzbar gewesen und die mangelnde Formatierung der Inhaltsverzeichnisse ist bis heute ein Nachteil für den Leser am Bildschirm. intelligentCAPTURE kann JEDES Buch bearbeiten mit Ergebnissen wie gedruckt; der Ansatz von Serial Solutions greift bei Neuerwerbungen oder Neuauflagen, welche eine ISBN erhalten (oder eingekaufte Daten).

GBV-Jahresbericht 2012 über 2011 – noch unveröffentlicht: Die kompletteste und aktuellste Sammlung von Inhaltsverzeichnissen hat der GBV mit 2.8 Millionen (Stand 31.12.2011), von denen 1,9 Mio mit konkreten Beständen der Bibliotheken verbunden sind. Es sind 30,7 % aller Monografien ab 1990 durch ein Inhaltsverzeichnis suchbar. Daran hat unser größter Mitbewerber Syndetics einen Anteil von 1,1 Millionen, unser Anteil war zu dieser Zeit erstmals darüber. Inzwischen liegt unser Anteil aus intelligentCAPTURE bei 1,5 Millionen (Stand Anfang September 2012 mit dandelon.com-Community mit DNB-Produktion und Casalini Libri).

Während auf die im GBV produzierten Daten im Schnitt 23 x pro Jahr zugegriffen wird, spielen die externen Lieferungen in der Nutzung eine wesentlich geringere Rolle (1,8 x für Serial Solutions). Während z.B. die ZBW alle ihre Neuerwerbungen bearbeitet und 35 Nutzungen pro Buch und Jahr misst, fallen Bibliotheken die wenig selbst beisteuern, deutlich zurück, partizipieren aber dennoch erheblich über den Verbundaustausch durch die Produktion der anderen.

Die Nutzung der TOCs nahm im GBV von 2010 auf 2011 um 35,1 % auf 15,6 Millionen Zugriffe zu. Dabei fanden 73,2 % der Zugriffe auf die im GBV produzierten Daten statt, obwohl deren Anteil nur bei 22,4 % liegt. Die eigene Produktion führt zur höchsten Nachfrage, die zugekauften Lizenzprodukte spielen im GBV eine untergeordnete Rolle. Die meisten produzierenden GBV-Bibliotheken nutzen intelligentCAPTURE.

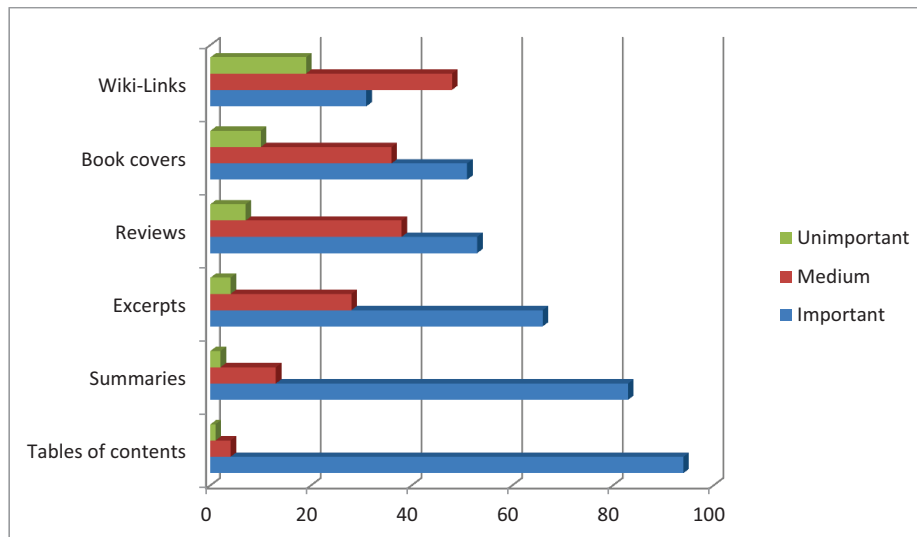
Eine vergleichbare Nutzungssteigerung misst die DNB, welche seit 2008 Neuerwerbungen und retrospektiv Inhaltsverzeichnisse produzieren lässt und über die Verbünde bereitstellt. Bis April 2012, also bis zur GND-Umstellung, sind hohe Steigerungen zu messen und eher unerwartet geht die Nutzung im Mai und Juni zurück und steigt erst wieder im Juli über das Vorjahr. In dieser kritischen Phase wurden wochenlang keine neuen Inhaltsverzeichnisse eingespielt, obwohl täglich produziert und bereitgestellt – eine vorübergehende Folgewirkung der GND-Umstellung und technischer Anpassungen. Spannend wird der Herbst, das Wintersemester. Es wäre verwunderlich, wenn ab Oktober die Nutzung sich grundsätzlich anders als die ersten vier Monate entwickeln würde – deutlich expansiv.



Die Nutzungssteigerungen beim GBV und DNB passen sehr gut zu den Benutzerbefragungen im Rahmen einer Studie mit 22.000 Bibliotheksbenutzern im HeBIS-Verbund aus dem Jahr 2011, 10 % aller registrierten Benutzer der HeBIS-Bibliotheken. In Hessen stammen die selbst produzierten Inhaltsverzeichnisse aus intelligentCAPTURE (UB Mainz, ULB Darmstadt, UB Frankfurt, UB Giessen, Herder-Institut Marburg mit integriertem dandelon.com-Austausch) bzw. kommen über den Verbundaustausch. 2004 war die Hans-Georg Nolte-Fischer, Direktor der ULB Darmstadt, Pionier unter allen deutschen Universitätsbibliotheken und produziert mit dem Ziel, den gesamten Freihandbereich von fast 1 Million Titel in der bald zu beziehenden neuen Bibliothek über Inhaltsverzeichnisse suchbar zu machen. Alternativ müssten die Benutzer zu einem Fitnesslauf zwischen Regalen und Stockwerken motiviert werden. Freihandaufstellungen allein lösen also keine Erschließungsprobleme, bieten aber Inhaltsverzeichnisse für noch nicht digital bearbeitete Titel und natürlich den kompletten Text in bester Lesequalität.

Die Studie zeigt, dass 94 % aller Benutzer im OPAC oder entsprechenden Suchsystem Inhaltsverzeichnisse nutzen, sprich suchen und anschauen, wollen. Danach folgen Zusammenfassungen, Textauszüge und Buchbesprechungen – also

ein Content, der über die bibliothekarische Kurzbeschreibung mit 3-5 Klassen oder Deskriptoren grundsätzlich deutlich hinausgeht. Nutzungswunsch und Nutzungsstatistik stimmen überein – keine Selbstverständlichkeit:



## 5. Weitere Entwicklungsziele

Die Erschließung neuer Regionen oder Zielgruppen (Expansionsstrategie) für die gleiche Technologie als auch die Entwicklung neuer Leistungen für die bestehenden Kunden (Diversifikationsstrategie) sind die zwei naheliegenden Optionen im Marketing-Latein. Beide Strategien kosten viel Geld, Zeit und Mühe. Lassen wir das Thema Internationalisierung hier in Jülich weg und schauen auf technologische Entwicklungen.

Ob man im Schnitt 2,5 Seiten für ein Inhaltsverzeichnis scannt, manche sind auch über 30 Seiten lang oder gleich ganze Aufsätze und Kapitel ist für Scanner und Rechner egal, die Limitierung hatte ja juristische Gründe, verbessern aber auch neben Recall die Precision der Retrieval-Lösungen. Ob man die Images selbst lokal generiert oder gleich ganze Laufwerke übernimmt, ist für Software intelligentCAPTURE eher äußerlich, der Workflow zur Verarbeitung setzt eben später

ein, zusätzliche Importprogramme sind rasch ergänzt. Die über 500.000 Zeitungsseiten von eliechtensteinensia.li sind übernommen von einem Spezialisten für die Konvertierung von Mikrofilm zu TIFF. Schwieriger ist jedoch die Zusammenführung mit den Bibliothekssystemen. Für die nächsten Jahre stehen starke Automatisierungen für die Titelaufnahme auf Basis von automatischen Dokumentstrukturanalysen an. Dafür musste die maschinelle Indexierung für sehr viele Sprachen ausgebaut werden und ist noch lange nicht am Ende. Ziel der langfristigen Entwicklung ist das Verständnis von komplexen Strukturen bis hin zu Aussagen, welche maschinelle Verfahren wie „Watson“ von IBM zu Antwortmaschinen machen können. Der Weg ist noch sehr lang. Wir als kleiner Player können dabei Nischen zu erobern versuchen.

Anstelle einer Literaturliste: <http://www.agi-imc.de> – Publikationen

## Netzwerkarbeit in der Fraunhofer-Gesellschaft

Andrea Siegborg

Abteilung Informationsmanagement, Fraunhofer Zentrale, Sankt Augustin

### Wissen Managen

Wir identifizieren für unsere Arbeit verschiedenste Wissensquellen, erwerben Wissen, bewahren und nutzen es und entwickeln daraus wieder neues Wissen. Dies geschieht zum Teil in den Köpfen und zum Teil in persönlichen Aufzeichnungen. Schwieriger wird es, wenn mehrere Personen diesen Prozess gemeinsam durchlaufen sollen. Hier reicht es nicht mehr, irgendwo irgendetwas zu sammeln oder abzulegen. Vielmehr muss ein geordneter Managementprozess für die Gemeinschaft etabliert werden, der jedem Einzelnen alle Schritte ermöglicht und die Ergebnisse für die Gemeinschaft zur Verfügung stellt.

Der Wissensmanagement-Kreislauf ( siehe Abb.1) beschreibt sechs operative Tätigkeiten und Managementaufgaben, die einen bestmöglichen und systematischen Umgang mit der Ressource Wissen ermöglichen.

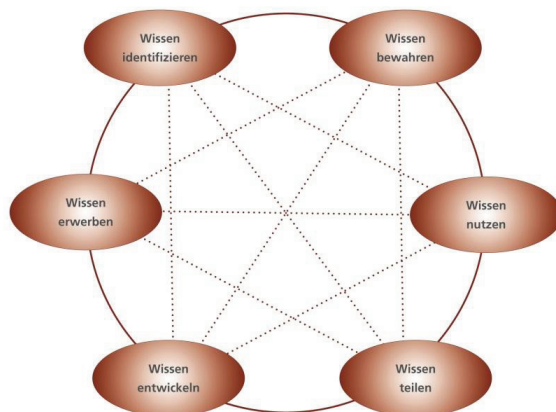


Abb. 1: Bausteinmodell des Wissensmanagements [Probst, Raub, Romhardt 1999]



Wissen identifizieren: Transparenz über alle Wissensquellen schaffen

Wissen erwerben: Steigerung der eigenen Wissensbasis mit Hilfe externer Wissensquellen

Wissen entwickeln: Aufbau neuer Fähigkeiten, Ideen und Prozesse

Wissen teilen: Bereitstellen des Wissens für die Mitarbeitenden

Wissen nutzen: Vorhandenes Wissen in laufende Aktivitäten einbinden

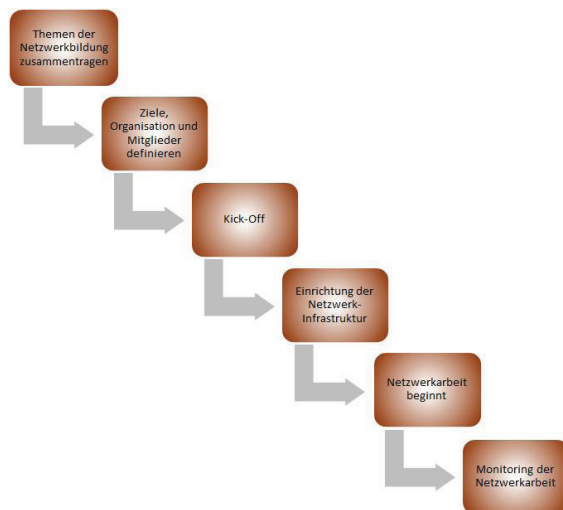
Wissen bewahren: Selektion von bewahrungswürdigem Wissen, angemessene Speicherung, regelmäßige Aktualisierung.

#### **Netzwerke bilden – Unterstützung von Anfang an**

Netzwerke sind kein Selbstzweck und funktionieren nicht von allein. Es sind Gemeinschaften, die sich zusammenschließen und gemeinsam an fachlichen Themen oder an einem gemeinsamen Projekt arbeiten. Alle Formen der Bereichs- und Instituts-übergreifenden Zusammenarbeit, in denen Wissen für ein gemeinsames Ziel oder für gemeinsame Interessen zwischen den beteiligten Experten ausgetauscht wird, nennt man Wissensnetzwerke. Die flexible Art des Zusammenschlusses bietet Kooperationsmöglichkeiten für strategische Allianzen, Möglichkeiten des Zusammenführens verschiedener Kompetenzen, aber auch Austausch- und Diskussionsmöglichkeiten für die Netzwerkpartner.

Erfolgreiche Netzwerke zeichnen sich durch ein anspruchsvolles Netzwerkmanagement aus. Dies beginnt schon bei der Auswahl der Partner, die die Ziele des Netzwerkes gemeinsam erreichen sollen. Neben dem gemeinsamen Netzwerkziel müssen natürlich auch die unterschiedlichen Interessen der einzelnen Partner berücksichtigt werden. Jeder soll schließlich einen Nutzen erhalten und sich dann auch zu den Verhaltensregeln der Gemeinschaft bekennen. Nicht zu vergessen ist auch die netzwerktragende Infrastruktur. Sie muss gut geplant, aufgebaut und kontinuierlich gepflegt werden.

Trotz guten Willens ist aber die Gefahr des Scheiterns einer Netzwerkkooperation groß, denn Aufbau und Betrieb sind anspruchsvolle Management-Aufgaben. Um allen Instituten den nachhaltigen Aufbau und den erfolgreichen Betrieb von Wissensnetzwerken zu ermöglichen, bietet die Abteilung Informationsmanagement der Fraunhofer-Gesellschaft Unterstützungs- und Beratungsleistungen für die Initiierung, den Aufbau und die Begleitung von Netzwerken an (Abb.2). Mit diesem Service werden Netzwerke bei allen relevanten und kritischen Schritten betreut, um die gewünschten Netzwerkziele effizienter zu erreichen.



**Abb. 2: Unterstützungsleistungen der Netzwerkbetreuung**

### **Netzwerke in der Fraunhofer-Gesellschaft**

Die Institute der Fraunhofer-Gesellschaft arbeiten unterschiedlich zusammen. Sie kooperieren in Verbünden, Allianzen und Innovationsclustern. Oder sie bündeln, je nach Anforderung, unterschiedliche Kompetenzen in flexiblen Netzwerken oder Communities mit dem Ziel, ihre führende Stellung bei der Entwicklung von Systemlösungen und der Umsetzung ganzheitlicher Innovationen zu sichern.

In Netzwerken können die Mitarbeitenden der Fraunhofer-Gesellschaft Informationen austauschen und ihre Erfahrungen einbringen und so von dem »gemeinsamen Gedächtnis« profitieren. Das Wissen ist nicht mehr isoliert über verschiedene

Abteilungen und Institute hinweg verteilt, sondern wird in einem Netzwerk gebündelt. Wissen und Erfahrungen aus den unterschiedlichen Forschungsbereichen der Fraunhofer-Gesellschaft sind so im Netzwerk verfügbar.

Fachlich verwandte Institute organisieren sich in Forschungsverbünden und treten gemeinsam am FuE-Markt auf. In der Fraunhofer-Gesellschaft gibt es folgende Verbünde: IuK-Technologie, Life Sciences, Mikroelektronik, Oberflächentechnik und Photonik, Produktion, Verteidigungs- und Sicherheitsforschung sowie Werkstoffe-Bauteile-Materials.

Institute mit unterschiedlichen Kompetenzen kooperieren in Fraunhofer-Allianzen, um ein Geschäftsfeld gemeinsam zu bearbeiten und zu vermarkten.

Eine weitere Form der Zusammenarbeit gibt es auf Projektebene. Weniger formal und zum Teil sehr dynamisch findet die Zusammenarbeit in den Netzwerken und Communities statt. Netzwerke und Communities sollen vor allem eins sein: Wissensnetzwerke für den persönlichen Austausch und zur schnellen Beantwortung konkreter Fragen. Durch den Wissenstausch können Best Practices etabliert werden, die wiederum von den Netzwerkteilnehmern übernommen werden. So existiert eine Vielzahl an Netzwerken, die bereits seit vielen Jahren verteilt zusammenarbeiten und ihr Wissen teilen.

Für eine gute Zusammenarbeit benötigt jedes Netzwerk den Zugriff auf die gemeinsame Wissensbasis. Dafür ist eine leistungsfähige Informations- und Kommunikationstechnik erforderlich. Die Abteilung Informationsmanagement bietet hierfür eine Kollaborations-Plattform an, die im Mitarbeiterportal der Fraunhofer-Gesellschaft integriert ist. Der Zugang zum Mitarbeiterportal erfolgt über die individuelle Smartcard und bietet einen sicheren Zugang zur persönlichen Arbeitsplattform. In der Kollaborations-Plattform selbst stehen nach dem einmaligen Login verschiedene Anwendungen zur Verfügung und ermöglichen eine unkomplizierte Zusammenarbeit.

### **Mitarbeiterportal – Einfach gut vernetzt**

Die »Kollaborations-Plattform« im Mitarbeiterportal der Fraunhofer-Gesellschaft wurde speziell für die verteilte Zusammenarbeit entwickelt. Sie ist der Startpunkt für die softwaretechnische Unterstützung des Wissensmanagements in der Fraunhofer-Gesellschaft.

Wissensmanagement ohne IT ist denkbar, doch es wird vor allem durch das Mitwirken der Mitarbeitenden gelebt. Teams, die aber an verschiedenen Orten arbeiten, kommen ohne Informations- und Kommunikationsanwendungen kaum weiter. Komplexe Projekte mit hoher Detailtiefe lassen sich kaum ohne Softwareunterstützung ordnen und organisieren. Die technische Ebene erhält so neben der personellen, methodischen und organisatorischen einen wichtigen Stellenwert.

Außer der Unterstützung beim Netzwerkaufbau hat die Abteilung Informationsmanagement im Mitarbeiterportal ein erstes Angebot zur technischen Netzwerkunterstützung aufgebaut. Dieses Angebot wird laufend an die Bedürfnisse der Netzwerke angepasst und technisch kontinuierlich erweitert.

Die Plattform ist einfach und intuitiv zu bedienen. Anwender greifen direkt über den Web-Browser mit ihrer Smart Card auf ihr personalisiertes Mitarbeiterportal zu. Über den Reiter »Kollaboration« gelangt man direkt in den Kollaborationsbereich.

Die Plattform lässt sich inhaltlich nach den individuellen Wünschen einrichten und bietet wichtige Tools für den Wissenstausch und die Zusammenarbeit an. Darüber hinaus werden dort verschiedene Informationskanäle und Services bereitgestellt.

## Livelihood – Der Wissenspeicher im Netzwerk

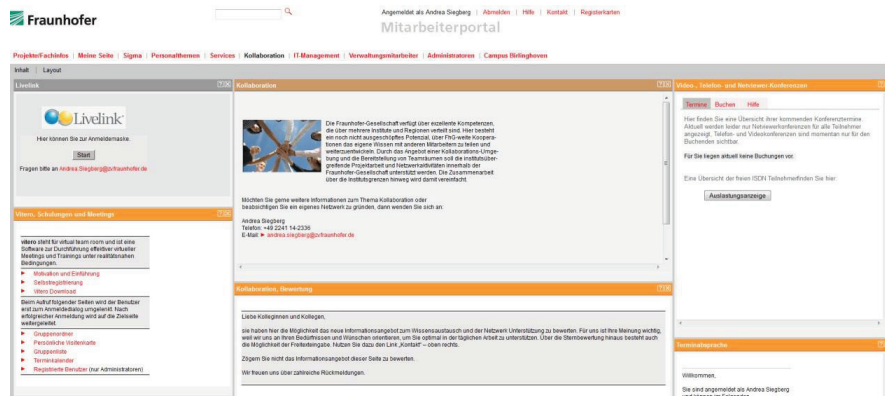


Abb. 3: Livelink im Mitarbeiterportal

Livelihood (Abb. 3) ist ein webbasiertes System, das die Zusammenarbeit in Netzwerken mittels Dokumentenverwaltung, Teamräume, Diskussionsforen, u.a.m. unterstützt. Das System ist im Kollaborationsbereich des Mitarbeiterportals eingebunden. Der registrierte Livelinknutzer kann auf Quellen zugreifen, öffnen, bearbeiten und wieder ablegen. Über ein umfangreiches Berechtigungskonzept können die Netzwerkmitglieder entscheiden, wer auf welche Inhalte und damit auf den Wissenspeicher des Netzwerkes zugreifen darf. Durch den kontinuierlichen Austausch in der kollaborativen Arbeitsumgebung wird das gemeinsame Wissen ständig gepflegt und erweitert. Darüber hinaus können die Netzwerkmitglieder sich in Foren austauschen und zu relevanten Themen diskutieren.

## Terminabsprache – Einfach und schnell Termine planen

Um einen gemeinsamen Termin zu finden, bedarf es meistens zahlreicher E-Mails und mehrerer Telefonate. Möchte man unkompliziert einen Termin für gemeinsame Sitzungen abstimmen, bietet sich das Tool Terminabsprache (Abb. 4) an. In sechs Schritten kann man Kolleginnen und Kollegen eine Auswahl an Terminen vorschlagen. Mit einer Antwort-Funktion kann jeder einzelne seine Präferenz

zurückmelden. So erhält der Organisator einen schnellen und vollständigen Überblick für die Verfügbarkeit der Teilnehmer und kann entsprechend planen.

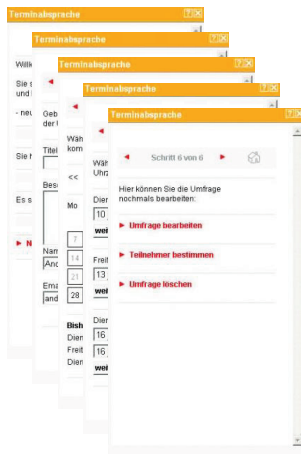


Abb. 4: In wenigen Schritten zum gemeinsamen Termin

### Vitero – Der virtuelle Teamraum

Vitero (Abb. 5) ist eine Kommunikationsplattform, die verteilte Netzwerkarbeit unterstützt. Der Einsatz ermöglicht eine Optimierung der Kommunikations- und Meeting-Kultur durch ortsunabhängige, einfach organisierbare und interaktiv gestaltbare virtuelle Sitzungen, Schulungen und Workshops. Das modulare System besitzt ein Content-Management-System sowie ein Benutzer-Management-System. Dokumente wie etwa Präsentationen oder Textdokumente können schon vor dem Meeting durch den Moderator für alle Teilnehmer verfügbar gemacht werden. Das Tool ist die erste Wahl, wenn komplexere Besprechungen in großen Gruppen durchgeführt werden sollen. Zusätzlich kann jede Sitzung durch einen Moderator und einen Co-Moderator geleitet werden.

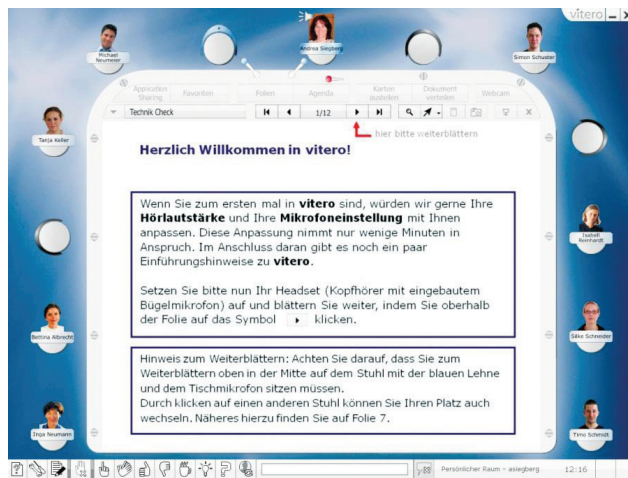


Abb. 5: Virtueller Teamraum in Viteo

## Netviewer – Konferenzen einfach und schnell

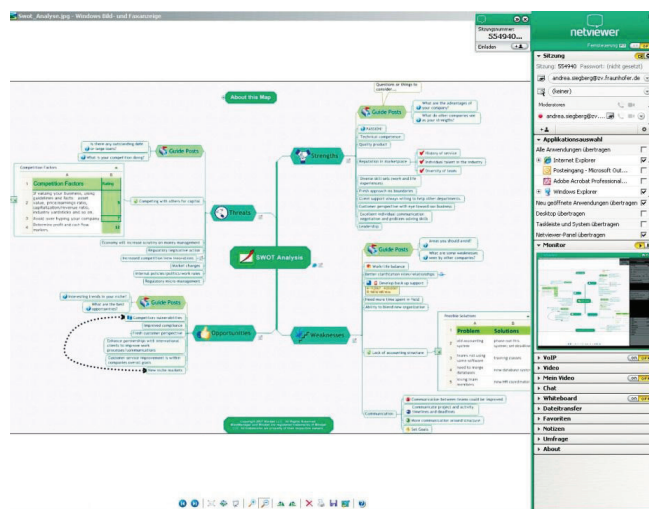


Abb. 6: Application Sharing in Netviewer

Über Netviewer (Abb. 6) können schnell und unkompliziert Webkonferenzen durchgeführt werden. Ein Doppelklick genügt, und die Teilnehmer kommen ganz einfach ohne Installation oder Änderungen an der Firewall über einen E-Mail-Link

zusammen. Die integrierte VoIP-Funktion ersetzt die Telefonkonferenz, und per Webcam können sich die Mitarbeiter gegenseitig sehen. Das Kommunikationsspektrum reicht von der klassischen Telefonkonferenz über die Videokonferenz bis hin zur modernen Web-Konferenz. Dabei stehen dem Nutzer neben der Kommunikationsanbindung Features wie Whiteboard, Dateitransfer, Chat, Desktop-Sharing oder ein Präsentationsmodus zur Verfügung.

### Fraunhofer-Wikifarm

Ein neuer Dienst für alle Mitarbeitenden der Fraunhofer-Gesellschaft ist die Fraunhofer-Wikifarm ( Abb. 7).



Abb. 7: Hauptseite Fraunhofer-Wikifarm

Auf der Wikifarm können sich Mitarbeitende unaufwändig und unbürokratisch ein Wiki auf einem zentralen Web-Server einrichten lassen. Mit Wikis lassen sich schnell und unkompliziert Inhalte kollaborativ erstellen und publizieren – ein Konzept das großen Anklang findet. Die Wikis können zur Informationsverbreitung, zur internen Dokumentation und Kommunikation und zur Unterstützung der Projektarbeit eingesetzt werden.

Auf einer zentralen Web-Plattform gibt es ein Masterwiki, in dem bereits Einstellungen und Funktionalitäten gespeichert sind, z.B. die zentralen Programmdateien, das Layout Template und diverse Extensionen. Die einzelnen Wikis greifen auf die zentralen Dateien zu und „erben“ die Einstellungen und Funktionalitäten des Masterwikis. Updates werden lediglich im



Masterwiki vorgenommen und wirken sich damit auf alle angegliederten Wikis aus. Den einzelnen Wikis auf der Fraunhofer-Wikifarm sieht man ihre Abhängigkeiten nicht direkt an – jedes hat seine eigene URL, eine eigene Nutzer- und Zugriffsverwaltung sowie lokale Einstellungen. Auch alle angelegten Wiki-Seiten werden pro Wiki gespeichert.

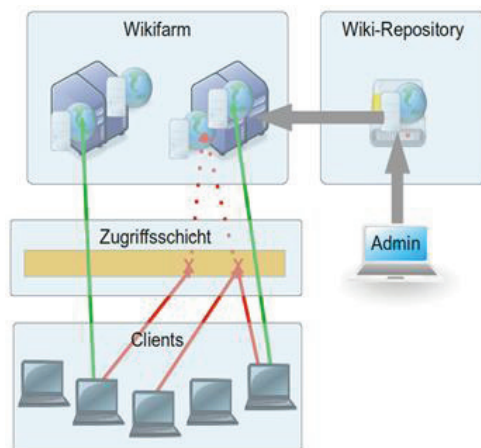
Für die einzelnen Wikis gibt es unterschiedliche Nutzungsrechte:

- Lesen: Fraunhofer- weit und Schreiben: Fraunhofer-weit (Kollaborative Wikis)
- Lesen: Fraunhofer-weit und Schreiben: nur Projektgruppe (Informationsverbreitung)
- Lesen: nur Projektgruppe und Schreiben: nur Projektgruppe (geschlossene Projektwikis).

Die Projektwikis können sowohl im Intranet, für die interne Projektarbeit als auch auf einem separaten Server, zur Nutzung mit externen Partnern, erstellt werden.

Die Wikiadministration erfolgt über eine Webmin-Administrationsumgebung, die auf die Anforderungen der Wikifarm angepasst ist. Über diese Administrationsumgebung werden die Wikis und zugehörige Datenbanken eingerichtet und angepasst / modifiziert (Abb. 8).

Ein Zugang zum Wiki ist sowohl über Passwort-Eingabe als auch über Smartcard möglich.



**Abb. 8: Administrationsumgebung der Fraunhofer-Wikifarm**

Die Fraunhofer Wikifarm wird auf Basis der Media Wiki-Technologie weiterentwickelt.

Aktuell bieten wir die Version 1.18 an.

Vor jedem Releasewechsel werden alle eingebunden Extensionen auf ihre Lauffähigkeit und Kompatibilität in der neuen Version getestet.

Nachdem die Tests erfolgreich abgeschlossen und dokumentiert wurden erfolgt die Umstellung auf die neue Release in allen Wikis.

## **Ausblick**

Aktuell wird in der Fraunhofer-Gesellschaft ein neues Mitarbeiterportal realisiert.

Die wesentlichen Zielsetzungen des Mitarbeiterportals sind dabei die folgenden Punkte:

- **Zentrale Portalplattform:** Zentrales, rollenbasiertes Portal für die verschiedenen Webdienste, Anwendungen und Informationen in der Fraunhofer-Gesellschaft. Zusätzlich soll damit den Instituten eine technisch und inhaltlich umfassende und attraktive Plattform angeboten werden, auf der sie eigene Institutsportale aufsetzen können.
- **Effektiver Wissenszugriff:** Die bestehenden Angebote und Inhalte sollen grundlegend überarbeitet werden, so dass eine fachlich sinnvolle und nutzerzentrierte Strukturierung entsteht, z.B. durch das Angebot von Fachportalen.
- **Elektronische Prozessunterstützung:** Die Mitarbeitenden sollen durch Workflows und Self-Services unterstützt werden. Dafür bereits bestehende Anwendungen sollen so weit wie möglich in das Portal integriert und angebunden werden.
- **Verbesserung der internen Kommunikation und Zusammenarbeit:** Die Kooperationsmöglichkeiten innerhalb der Fraunhofer-Gesellschaft sollen durch innovative, in das System integrierte Kooperationswerkzeuge erweitert werden.

---

## **Vernetzung biographischer Online-Ressourcen – Catalogus Professorum Rostochiensium und Rostocker Matrikelportal**

Robert Stephan, Karsten Labahn, Matthias Glasow

Universität Rostock

### **Zusammenfassung**

In den beiden universitätsgeschichtlichen Projekten „Catalogus Professorum Rostochiensium“ und „Rostocker Matrikelportal“ werden Personendaten zu Professoren und Studierenden erhoben, aufbereitet und präsentiert. Durch die Anreicherung der Datenbestände mit weiteren relevanten Ressourcen und die Vernetzung mit anderen Internetangeboten können attraktive biographische Informationssysteme aufgebaut werden. Zum einen werden konsequent und in großer Zahl Digitalisate von Originalquellen, Bildern, Archivmaterialien und anderen Dokumenten zielgenau verlinkt oder direkt in die Einträge eingebunden. Zum anderen werden über die Verwendung von Personennormdaten der Gemeinsamen Normdatei (GND), des Austauschformats „Beacon“ sowie des BEACON-AKS-Webservice tausende von Verknüpfungen zu einer Vielzahl von ganz verschiedenartigen weiteren Internetangeboten automatisch generiert.

### **Abstract**

In the projects “Catalogus Professorum Rostochiensium” (Catalogue of Professors of Rostock University) and “Rostocker Matrikelportal” (Matriculation Registers of Rostock University) historical biographical information on professors and students is collected, prepared and presented in the internet. By enriching the data with different relevant documents and by crosslinking with other online-resources attractive biographical information systems can be created. On the one hand a large number of digitized original sources, portraits, archival material and other documents are linked or directly integrated into the web presentation. On the other hand name authority file records of the “Gemeinsame Normdatei” (GND), the exchange format “Beacon” and the BEACON-AKS webservice are used to automatically generate thousands of links to a multitude of other biographical internet resources.

## Einleitung

Im Hinblick auf das 600jährige Jubiläum der Universität Rostock im Jahr 2019 widmen sich die Universitätsbibliothek und die Forschungsstelle Universitätsgeschichte der Universität Rostock in zwei Online-Datenbankprojekten der Erforschung der beiden wichtigsten Personengruppen, die die Institution Universität über die Jahrhunderte geprägt haben – den Professoren und den Studenten. Die Informationssysteme „Catalogus Professorum Rostochiensium“ und „Rostocker Matrikelportal“ sind geeignet, den Mehrwert einer konsequenten Vernetzung zwischen institutionellen Repositorien und personenbezogenen Forschungsdaten zu demonstrieren. Darüber hinaus sind beide Anwendungen über Normdaten mit einer wachsenden Zahl von anderen biographischen Internetangeboten vernetzt.

Im Folgenden werden zunächst das Matrikelportal, der Professorenkatalog und der Dokumentenserver der Universität Rostock kurz vorgestellt. Anschließend wird auf die Verknüpfung von Personendaten mit digitalisierten Dokumenten und schließlich auf die projektübergreifende Vernetzung auf der Grundlage von Personennormdaten eingegangen.

## Matrikelportal, Professorenkatalog und Dokumentenserver

Das *Rostocker Matrikelportal*<sup>1</sup> enthält im Kern Datenbanken mit allen Immatrikulationen an der Universität Rostock 1419-1945 – eine Fortführung bis in die Gegenwart ist geplant. Die Datengrundlage bilden die im Universitätsarchiv überlieferten Matrikelbücher, in denen die lange Reihe der Universitätsbesucher lückenlos dokumentiert ist. Die Eingabe der Matrikeleinträge erfolgt anhand von digitalisierten Vorlagen über ein Online-Formular. Einheitliche Übertragungsrichtlinien und Korrekturvorgänge gewährleisten die Einhaltung der etablierten Standards wissenschaftlicher Quelleneditionen. Zur Aufbereitung der Daten erfolgte eine Normalisierung und Zuordnung der oft stark unterschiedlichen Schreibformen von Orts- und Personennamen. Neben Zeitpunkt, gezahlten Gebühren und Bemerkungen zu den Umständen der Immatrikulation enthalten die Matrikeleinträge in der Regel den Namen, die Herkunft und Hinweise zum Status der eingeschriebenen Person (akademische Titel, Adelsprädikate, Ämter).

---

<sup>1</sup> <http://matrikel.uni-rostock.de>

In die tabellarischen Matrikelbücher des 19./20. Jahrhunderts schrieben sich die Studenten selbst ein und machten weitere Angaben beispielsweise zum Schulabschluss, Wohnort, Beruf des Vaters und Studienfach. Insgesamt sind im Portal zurzeit mehr als 80.000 Personenzugnisse zugänglich. Suchfunktionen ermöglichen eine Recherche nach Personennamen und Herkunftsorten, aber auch detaillierte Datenbankabfragen nach Studienfächern, akademischen Titeln, Geschlecht oder Religion. Neben den Matrikeldaten macht das Portal auch Scans der Originalquellen zugänglich. Die Anreicherung mit Geodaten ermöglicht die Visualisierung der Herkunftsorte der Studenten auf einer interaktiven Karte. Eine Nutzerkommentarfunktion bietet den Besuchern des Portals die Möglichkeit, alle Matrikeleinträge zu kommentieren. Für die Realisierung des Frontends zum Matrikelportal kommen Java-Webtechnologien (Java Server Pages) zum Einsatz. Die Daten selbst werden in einer MySQL-Datenbank gespeichert. Für die Suche wurde ein SOLR<sup>2</sup>-Index erstellt. Sogenannte Data Import Handler ermöglichen es, innerhalb des SOLR-Indexes mehrere Datenfelder aus verschiedenen Tabellen zu einem Suchfeld zusammenzufassen. Außerdem werden Features, wie Umlautnormalisierung, Stammwortreduktion, Stoppwortfilter verwendet.

Im *Catalogus Professorum Rostochiensium* (CPR)<sup>3</sup> sollen alle Professoren der Universität Rostock dokumentiert werden. Konzeptionell verwendet der CPR sowohl Elemente prosopographischer Datenbanken mit statistischen Auswertungsmöglichkeiten als auch biographischer Lexika mit ausformulierten Personenartikeln. Im Gegensatz zu meist alphabetisch nach den Namen sortierten Personenlexika weist der CPR eine inhaltliche Systematik auf: die Lehrstühle/ Professuren und ihre Inhaber in chronologischer Folge. Die Professoreneinträge werden mit dem Vorgänger und Nachfolger in der Professur (sofern vorhanden) verknüpft. Die Kerninformationen des Katalogs zu Art, Zeitraum, Fachgebiet der Rostocker Professuren und Dozenturen werden vollständig recherchiert und in strukturierter und einheitlicher Form gespeichert. Jeder Eintrag im Catalogus enthält mindestens diese Grundangaben (Kurzeinträge). Für den Zeitraum ab 1900 werden die Kerninformationen ergänzt mit aus den Quellen recherchierten bzw. auf eigenen Angaben beruhenden Biographien mit Angaben zu wissenschaftlichem Profil und Werdegang, zu Publikationen, Mitgliedschaften und Funktionen in Wissenschaft und

---

<sup>2</sup> <http://lucene.apache.org/solr>

<sup>3</sup> <http://cpr.uni-rostock.de>

Gesellschaft (Langeinträge). Für Professoren der früheren Zeit gibt es in der Regel bereits vielfältige biographische Literatur, die entweder verlinkt oder direkt in die Einträge eingebunden wird (Dokumenteinträge). Für die Erfassung, Bearbeitung, Präsentation und Recherche der biographischen Daten wird MyCoRe<sup>4</sup> eingesetzt, eine Software zum Aufbau von Dokumenten- und Publikationsservern, Archivanwendungen oder digitalen Sammlungen. MyCoRe bietet einen gemeinsamen Kern, der die typischerweise für solche Anwendungen erforderliche Funktionalität, wie z. B. Metadatenverwaltung, Suchfunktionen oder OAI-Schnittstelle bereitstellt. Für den Aufbau eines Datenmodells stehen eine Vielzahl von Datentypen zur Verfügung sowie zusätzlich die Möglichkeit, beliebiges XML in den Metadaten zu speichern. Damit lassen sich die umfangreichen biographischen Informationen zu Rostocker Professoren gut abbilden. LexM, das Lexikon verfolgter Musiker und Musikerinnen der NS-Zeit<sup>5</sup>, ist ein weiteres MyCoRe-basiertes biographisches Informationssystem.

Die UB Rostock ist Mitglied der Entwickler-Community und betreibt auf Basis von MyCoRe neben der Spezialanwendung Catalogus Professorum vor allem Dokumentenserver für die Hochschule Neubrandenburg, für die Landesbibliothek Mecklenburg-Vorpommern und für die Universität Rostock. Der *Rostocker Dokumentenserver* (RosDok)<sup>6</sup> wird einerseits als Hochschulschriftenserver für die Open-Access-Publikation von Dissertationen und universitären Reihen, andererseits aber auch für die Veröffentlichung der Ergebnisse aus Digitalisierungsprojekten eingesetzt.

### **Vernetzung von Personendaten und digitalisierten Dokumenten**

Die Einträge im Matrikelportal und im Professorenkatalog werden konsequent mit weiteren relevanten Ressourcen angereichert. So ist jeder Datensatz der Matrikeldatenbank mit Scans der Originalquellen verknüpft. Einer oder mehrere Links führen direkt auf digitale Faksimiles der entsprechenden Seiten der handschriftlichen Matrikelbücher bzw. der Matrikeledition von Adolf Hofmeister/Ernst Schäfer (um 1900) auf dem Rostocker Dokumentenserver. Zusätzlich wird von jeder Immatrikulation auf Digitalisate der historischen Vorlesungsverzeichnisse und auf

---

<sup>4</sup> <http://www.mycore.org>

<sup>5</sup> <http://www.lexm.uni-hamburg.de>

<sup>6</sup> <http://rosdok.uni-rostock.de>

den Professorenkatalog verwiesen, sodass Informationen zu den im Semester der Einschreibung angebotenen Vorlesungen sowie zu den an der jeweiligen Fakultät tätigen Professoren direkt verfügbar sind. Insgesamt erfolgt die Vernetzung von Matrikelportal und Dokumentenserver derzeit über etwa 180.000 automatisch generierte Links.

The screenshot displays the 'Catalogus Professorum Rostochiensium' website. The header includes the University of Rostock logo and navigation links. The main content area shows the profile of Paul-Friedrich Scheel, Prof. Dr. med. habil. The profile includes a list of his positions from 1939 to 1957, a list of attached documents (Anhang) such as a bibliography, curriculum vitae, and NSDAP experiences, and a list of images (Bild) including a portrait and a signature. The footer of the profile section provides the date of entry/modification and the recommended citation format.

**Universitt Rostock** Traditio et Innovatio

Start | Sitemap | Impressum | [Anmelden]

Catalogus Professorum Rostochiensium

Suchen » Dokumenten-Details

Katalogeintrag Biographie Dokumente/Anhang Langeintrag

**Scheel, Paul-Friedrich**  
Prof. Dr. med. habil.

Prof. in Rostock:

1939-1945	Dozent fr Orthopdie
1950-1951	Dozent fr Orthopdie
1951-1954	Professor mit Lehrauftrag fr Orthopdie
1954-1957	Professor mit Lehrstuhl fr Orthopdie

Anhang:

- Schriftenverzeichnis (1949, Personalakte, UAR)  
sheel\_paul-friedrich\_bib.pdf (258,2 KB)
- Lebenslauf (1947, Personalakte, UAR)  
sheel\_paul-friedrich\_cv.pdf (602 KB)
- Erfahrungen in der NSDAP und Einstellung zum neuen Deutschland (1946, Personalakte, UAR)  
sheel\_paul-friedrich\_erfahrungen\_nsdap.pdf (2,42 MB)
- Dokument (Biographischer Artikel von Eileen Stammer, 2010)  
sheel\_paul-friedrich\_bio.html (6,9 KB)
- Bild (Foto, Jahr unbekannt, UAR)  
sheel\_paul-friedrich\_pic.jpg (46,4 KB)
- Bild (Unterschrift)  
sheel\_paul-friedrich\_sig.jpg (23 KB)

eingestellt / gendert:  
23.02.2010, mglasow / 24.04.2012, mglasow

Empfohlene Zitierweise:  
Eintrag von "Paul-Friedrich Scheel" im Catalogus Professorum Rostochiensium,  
URL: [http://cpr.uni-rostock.de/metadata/cpr\\_person\\_00002680](http://cpr.uni-rostock.de/metadata/cpr_person_00002680)  
(abgerufen am 16.08.2012)

### Dokumentanhang im Rostocker Professorenkatalog

Auch im Catalogus Professorum werden Links zu RosDok generiert, ber die seitengenau auf einzelne Nachweise zu einem Professor, etwa in der Gelehrtenzeitschrift „Etwas von gelehrten Rostockschen Sachen“ und weiterer



digitalisierter biographischer Literatur, verwiesen wird. Vor allem aber wird die Repository-Funktionalität von MyCoRe zum Speichern und Verwalten von Dokumenten im Rahmen des Professorenkatalogs genutzt, um die Katalogeinträge mit weiteren Quellen, wie Portraits, Literaturverzeichnissen, Briefen, Archivmaterialien und anderen Dokumenten anzureichern. Hierbei werden bereits bestehende Biographien zu Professoren der älteren Zeit (vor 1900) sowie neu verfasste biographische Artikel direkt im CPR-Eintrag angezeigt (Reiter „Biographie“). Weitere Digitalisate werden im Dokumentanhang bereitgestellt. Die folgende Aufstellung gibt einen Überblick zu den im Professorenkatalog zugänglichen weiteren Dokumenten:

Bilder (Portraits)	1311
Unterschriften	363
biographische Artikel (Volltext)	120
Lebensläufe, Schriftenverzeichnisse (von Professoren selbst zur Verfügung gestellt)	1224
Digitalisate biographischer Literatur	484
Digitalisate aus Archivalien (Personalakten)	503
Sonstiges	18
Gesamt	4023

**Dokumente in den Einträgen des Catalogus Professorum**

### **Personennormdaten als Grundlage für die projektübergreifende Vernetzung**

Der wohl größte Bestand an Personendatensätzen in Deutschland wird von der Deutschen Nationalbibliothek (DNB) in der Gemeinsamen Normdatei (GND)<sup>7</sup> verwaltet. In der GND wurden die bisherigen Normdateien PND (Personennamendatei), GKD (Körperschaftsdatei) und SWD (Schlagwortnormdatei) zusammengeführt. Die etwa sieben Mio. Personendatensätze der GND dienen vorrangig dazu, Autoren von Publikationen eindeutig zu identifizieren. Sie bieten

---

<sup>7</sup> [http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd\\_node.html](http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html)

jedoch auch eine gute Grundlage für die Vernetzung biographischer Ressourcen aus dem wissenschaftlichen Umfeld, da die meisten der behandelten Personen selbst publiziert haben oder über sie publiziert worden ist. Die Pflege der Normdatei erfolgt kooperativ. In jeder größeren Bibliothek gibt es geschultes Personal, welches Personendatensätze in der GND anlegen und erweitern kann. Deshalb empfiehlt es sich für biographische Projekte eine Bibliothek als Partner zu suchen. Über eine Webschnittstelle lassen sich einzelne Datensätze im RDF-Format abrufen und weiterverarbeiten. Außerdem bietet die DNB einen regelmäßig aktualisierten Komplettabzug zum Download und einen Update-Service an.

Alle ca. 2.000 Personen im Rostocker Professorenkatalog wurden mit der GND abgeglichen. In vielen Fällen wurden durch die Universitätsbibliothek Rostock neue GND-Datensätze angelegt. Oftmals konnten die Datensätze in der GND durch Angaben aus dem Professorenkatalog angereichert und Fehler in beiden Datenbeständen ermittelt und korrigiert werden. Im Matrikelportal wurden bisher ca. 3.200 Datensätze mit der GND verknüpft. Da viele Studierende mehrfach in den Matrikel Datenbanken vorkommen – z. B. aufgrund einer erneuten Immatrikulation nach dem Besuch einer anderen Universität oder eines zusätzlichen Eintrags zur Promotion in einem der Fakultätsbücher –, sind damit etwa 2.000 Personen identifiziert. Die GND-Verknüpfung erfolgt in der Regel nur, wenn ein Nutzer des Matrikelportals einen Kommentar zu einem Matrikeleintrag verfasst, der Hinweise auf den weiteren Lebensweg dieser Person gibt. Eine vollständige Aufnahme der gesamten Matrikel in die GND ist personell und technisch nicht zu leisten und auch nicht sinnvoll, da die Angaben in der Matrikel häufig nicht ausreichen, um eine Person eindeutig zu identifizieren.

### **BEACON – Datenformat und Infrastruktur für die Vernetzung**

Über das BEACON Format<sup>8</sup> lassen sich Links zu Webseiten realisieren, deren Inhalte mit Normdaten verknüpft sind. Es wird derzeit für Personendaten genutzt, ist jedoch prinzipiell auch für weitere Normdaten (Körperschaften, Schlagworte) geeignet. Es handelt sich um eine einfache Textdatei, die aus mehreren Header- und Datenzeilen besteht. Jede Datenzeile besteht in der Regel aus einer GND-Nummer.

---

<sup>8</sup> <http://de.wikipedia.org/wiki/Wikipedia:BEACON>

Sie kann um die Anzahl der bereitgestellten Ressourcen und einen Link zum Zugriff auf diese Ressourcen ergänzt werden. Das Headerfeld #PREFIX gibt an, wie die URI des GND-Datensatzes aufgebaut ist. Im Headerfeld #TARGET wird die Struktur des Links auf das eigene Webangebot spezifiziert. Weitere Headerfelder enthalten Angaben zum Format, zum Herausgeber der Daten und Informationen wann die Datei erstellt wurde und wie häufig mit einer Aktualisierung zu rechnen ist. Derzeit laufen Bestrebungen das Beacon-Format als Request For Comments (RFC) bei der Internet Engineering Task Force (IETF) zu standardisieren. Der Entwurf von Jakob Voß, Mathias Schindler und Thomas Berger steht bis Dezember 2012 zur Diskussion.<sup>9</sup>

```
#FORMAT: PND-BEACON
#VERSION: 0.1
#FEED: http://cpr.uni-rostock.de/cpr_pnd_beacon.txt
#TARGET: http://cpr.uni-rostock.de/pnd/{ID}
#PREFIX: http://d-nb.info/gnd/
#NAME: Catalogus Professorum Rostochiensium (CPR)
#DESCRIPTION: Im Catalogus sollen alle Professoren der Universität Rostock
seit 1419 dokumentiert werden (im Aufbau).
#CONTACT: Forschungsstelle Universitätsgeschichte Rostock
<unigeschichte@uni-rostock.de>
#INSTITUTION: Universitätsbibliothek Rostock
#ISIL: DE-28
#MESSAGE: Eintrag im Catalogus Professorum Rostochiensium (CPR)
#UPDATE: will be rebuilt on every request
#TIMESTAMP: 2012-08-05T13:48:38Z
#REVISIT: 2012-08-12T13:48:38Z
11584113X
121269450
130380954
140749721
136694233
10852485X
132697696
129228796
...
```

#### **Aufbau der Beacon-Datei des Rostocker Professorenkatalogs**

Die Webseite [beacon.findbuch.de](http://beacon.findbuch.de) stellt einen Dienst zu Verfügung, der zurzeit etwa 180 Beacon-Dateien aus verschiedensten Projekten aggregiert. Darunter sind Beacon-Dateien zu den Katalogen der Bibliotheksverbünde mit jeweils mehreren hunderttausend Einträgen, zu den Personenartikeln der Wikipedia, zu Bibliographien

---

<sup>9</sup> <http://gbv.github.com/beaconspec/beacon.html> .

und Nachlassverzeichnissen, Biographien und biographischen Lexika, universitätsgeschichtlichen Projekten, Editionen, Volltextsammlungen und Bilddatenbanken. Bei Übergabe einer GND-Nummer liefert der Dienst PND/BEACON-AKS<sup>10</sup> eine Liste mit Internetressourcen zurück, die Informationen zu der mit dieser Nummer identifizierten Person anbieten. Auf diese Weise werden im Catalogus Professorum und im Matrikelportal Links zu einer Vielzahl von verschiedenen externen Informationsangeboten zu einem Professor oder Studierenden automatisch generiert. Ebenso kann aus anderen Projekten heraus auf die Rostocker Matrikel- oder Professoreneinträge verwiesen werden.

**Immatrikulation von Georgius Christophorus Detharding**

Transkription	normierte Angaben
Semester: 1715 Mich.	
Nummer: 2	
Datum: ~10.1715	
Vorname: Georgius Christophorus	Georg Christoph
Nachname: Detharding	Detharding
Herkunft: Rostochiensis	Rostock

**Weitere Informationen**  
PND: 104103329

- Wikipedia-Personenartikel
- Biographien in der Allgemeinen Deutschen Biographie [Detharding, Georg Christoph]
- Biographische Informationen des Registers zu [www.deutsche-biographie.de](http://www.deutsche-biographie.de)
- Allgemeine Deutsche Biographie (Wikisource)
- Eintrag im Catalogus Professorum Rostochiensis (CPR)
- Rostocker Matrikelportal 1419-1945
- Bayerische Staatsbibliothek (24)
- 105 Titel im Verbundkatalog des GBV [143268635]
- HeBIS-Verbundzentrale (6)

**Quellen**  
Edition Hofmeister  
Matrikelbuch 1419-1760

**Professoren/Vorlesungen**  
Professoren im Semester der Einschreibung

**Weitere Informationen**  
PND: 104103329

- Wikipedia-Personenartikel
- Biographien in der Allgemeinen Deutschen Biographie [Detharding, Georg Christoph]
- Biographische Informationen des Registers zu [www.deutsche-biographie.de](http://www.deutsche-biographie.de)
- Allgemeine Deutsche Biographie (Wikisource)
- Eintrag im Catalogus Professorum Rostochiensis (CPR)
- Rostocker Matrikelportal 1419-1945
- Bayerische Staatsbibliothek (24)
- 105 Titel im Verbundkatalog des GBV [143268635]
- HeBIS-Verbundzentrale (6)

**Integration von Links zu weiteren Online-Ressourcen unter Verwendung des Dienstes [beacon.findbuch.de/seealso/pnd-aks](http://beacon.findbuch.de/seealso/pnd-aks)**

<sup>10</sup> <http://beacon.findbuch.de/seealso/pnd-aks>

Insgesamt ist das Matrikelportal nach aktuellem Stand über etwa 5.700 einzelne Links mit 97 verschiedenen Internetangeboten vernetzt. Allerdings sind bisher nur etwa 5 % der Matrikeleinträge mit GND-Nummern versehen, während dies im Professorenkatalog nahezu vollständig der Fall ist. Insgesamt fast 3.700 einzelne Links verweisen von Einträgen im Catalogus Professorum auf 81 externe Online-Ressourcen. Die verschiedenen Internetangebote sind dabei in sehr unterschiedlicher Häufigkeit vertreten. So ist beispielsweise nur eine Person sowohl im Rostocker Professorenkatalog als auch im Bayerischen Musiker-Lexikon Online (BMLO) oder in der Sächsischen Biographie vertreten. In der folgenden Übersicht sind diejenigen Internetangebote aufgeführt, deren „Schnittmenge“ mit dem Catalogus Professorum mehr als 20 Personen umfasst. Nicht berücksichtigt sind hierbei aufgrund ihrer großen Zahl die Verlinkungen zu Bibliothekskatalogen. Zu beachten ist, dass in vielen Projekten noch nicht alle behandelten Personen mit GND-Nummern identifiziert worden sind, sich die Anzahl der Verknüpfungen also laufend vergrößert.

Deutschsprachige Wikipedia	562
Biographische Informationen im Index zur Deutschen Biographie	456
Kalliope Nachlässe und Autographen	347
Rostocker Matrikelportal	319
Landesbibliographie M-V	258
Biographien der Allgemeinen Deutschen Biographie (ADB)	173
Rektoratsreden im 19. und 20. Jahrhundert	157
Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts (VD 17)	140
Biographien der Neuen Deutschen Biographie (NDB)	108
Leichenpredigten 1550-1800 (GESA)	89
Bildarchiv Foto Marburg	87
Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 16. Jahrhunderts (VD 16)	60
Biographisch-Bibliographisches Kirchenlexikon (BBKL)	57
Zentrale Datenbank Nachlässe	54
Mitglieder der Leopoldina	49
Mitglieder der Vorgängerakademien der Berlin-Brandenburg. Akademie der Wissenschaften	45
Eintrag in den biogr. Datenbanken "Wer war wer in der DDR?" und "Deutsche Kommunisten"	36
Autographennachweise in der SLUB Dresden	34
Personen im Informationssystem Universitätssammlungen in Deutschland	34

Biographisches Lexikon hervorragender Ärzte (Pagel, 1901)	33
Wikisource	33
Professorenkatalog der Universität Leipzig	28
Wikimedia Commons	27
Bibliographieportal zur Geschichte Ostmitteleuropas	26
Mitglieder der Bayerischen Akademie der Wissenschaften	25
Historische Vorlesungsverzeichnisse der Universität Leipzig	20

**Internetangebote mit der größten Zahl von Verlinkungen mit CPR-Einträgen  
über den Dienst [beacon.findbuch.de/seealso/pnd-aks](https://beacon.findbuch.de/seealso/pnd-aks)**

Die Verteilung der Verknüpfungen auf die einzelnen Catalogus-Einträge ist allerdings sehr unterschiedlich. So finden sich für mehr als die Hälfte der Professoren des 16. bis 19. Jahrhunderts Informationen in fünf oder mehr anderen biographischen Projekten, während für einen Großteil der nach 1945 berufenen und vor allem der aktuell tätigen Professoren über die Bibliothekskataloge hinaus bisher gar keine weiteren Nachweise geliefert werden. Dies erklärt sich zum einen daraus, dass die GND-Nummer zur einheitlichen Identifizierung von Personen bislang vor allem im bibliothekarischen Kontext und in historisch orientierten biographischen Projekten verwendet wird. Zum anderen wird hier deutlich, dass die biographische Aufarbeitung der Wissenschafts- und Universitätsgeschichte der jüngeren Zeit noch ein Desiderat darstellt. Ein Schwerpunkt des Rostocker Professorenkatalogs liegt deshalb in der Erstellung ausführlicher aus den Quellen und der Literatur erarbeiteter biographischer Einträge für die Professoren der DDR-Zeit.

## Fazit

Forschungsprojekte und Dokumentenserver können von der konsequenten Vernetzung ihrer Inhalte profitieren. Institutionelle Repositorien sind Dienstleister für die Forschung. Sie können dem Bedarf zur Entwicklung von passgenauen Einzellösungen für kleinere und spezielle Projekte entsprechen. Dokumentenserver stellen dabei digitale Ressourcen langfristig und persistent referenzierbar zur Verfügung. Durch die Möglichkeit, auch auf Teile innerhalb der Dokumente zu verlinken, lassen sich die relevanten Informationen direkt in die Portale integrieren. Andererseits können die im Rahmen von Forschungsprojekten erfassten Daten

helfen, die Qualität der Dokumente, z. B. durch die Anreicherung mit Strukturinformationen, zu verbessern.

Die Verknüpfung von Informationsressourcen zu derselben Person innerhalb einer Vielzahl von ganz verschiedenartigen Internetangeboten lässt sich über die Verwendung der GND-Nummer, des Beacon-Formats sowie des BEACON-AKS-Webservice sehr einfach realisieren. Der größte Arbeitsaufwand ist hierbei für die Anreicherung der eigenen personenbezogenen Daten mit der GND-Nummer als einheitlichem Identifier erforderlich. Der Grad der Vernetzung erreicht im Vergleich zur konventionellen biographischen und prosopographischen Forschungspraxis eine ganz neue Dimension durch die massenhafte automatische Generierung von zielgenauen Verknüpfungen zu möglicherweise nicht bekannten Datenbeständen. Automatisch erfolgt auch die Aktualisierung und Erweiterung der Verlinkungen, sodass eine manuelle Prüfung und Pflege nicht erforderlich ist.

In den beiden universitätsgeschichtlichen Projekten zu Studierenden und Professoren der Universität Rostock werden Personendaten erhoben, die für sich genommen nur recht begrenzt aussagekräftig sind. Erst die Anreicherung mit weiteren Ressourcen und die Vernetzung mit anderen Datenbeständen eröffnete die Möglichkeit, für die Nutzer attraktive biographische Informationssysteme zu schaffen. Die wissenschaftliche Bedeutung des Catalogus Professorum und des Matrikelportals ergibt sich vor allem durch eine Auswertung in der Gesamtschau eines vernetzten Raums geisteswissenschaftlicher Forschungsdaten.

## **Zur Entwicklung eines Linked-Open-Data-Dienstes für Bibliotheksdaten**

Felix Ostrowski<sup>1</sup>, Adrian Pohl<sup>2</sup>

<sup>1</sup> [literarymachine.net](http://literarymachine.net)

<sup>2</sup> Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, hbz

### **Zusammenfassung**

Das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) startete seinen Linked-Open-Data-Service [lobid.org](http://lobid.org) im August 2010 und verbessert seitdem kontinuierlich die dem Dienst zugrundeliegenden Konvertierungsprozesse, Datenmodelle und Software. Dieser Beitrag erläutert zunächst Hintergrund und Motivation für die Entwicklung von [lobid.org](http://lobid.org). Er beschreibt dann das zugrundeliegende Software-Framework *Phresnel*, das in PHP geschrieben ist und Präsentations- wie Editierungsmöglichkeiten von RDF-Daten auf Basis des Fresnel Display Vocabulary for RDF bereitstellt. Das Papier gibt einen Überblick über den derzeitigen Stand der *Phresnel*-Entwicklung und diskutiert die technischen Herausforderungen. Schließlich werden mögliche Aussichten für die weitere *Phresnel*-Entwicklung skizziert.

### **Abstract**

The Northrhine-Westphalian Library Service Center (hbz) launched its LOD service [lobid.org](http://lobid.org) in August 2010 and since then has been continuously improving the underlying conversion processes, data models and software. This paper first explains the background and motivation for developing [lobid.org](http://lobid.org). It then describes the underlying software framework *Phresnel* which is written in PHP and provides presentation and editing capabilities of RDF data based on the Fresnel Display Vocabulary for RDF. The paper gives an overview of the current state of *Phresnel* development and discusses the technical challenges encountered. Finally, possible prospects for further developing *Phresnel* are outlined.



## Einleitung

Das Thema Linked Open Data (LOD)<sup>1</sup> hat in den letzten Jahren zunehmende Aufmerksamkeit erlangt, und es lassen sich vermehrt entsprechende bibliothekarische Projekte ausmachen. Seit 2009 beschäftigt sich das hbz sowohl mit den rechtlichen als auch den technischen Aspekten von Linked Open Data. Im August 2010 startete das hbz seinen LOD-Service lobid.org – das Akronym steht für “Linking Open Bibliographic Data” – und verbessert diesen Dienst seitdem kontinuierlich.

Im Kontext von lobid.org wird auf eine Weiterentwicklung des zugrundeliegenden Software-Frameworks abgezielt, so dass es read/write-fähig wird, das heißt sowohl die HTML-Präsentation der RDF-Daten als auch die Generierung von Webformularen zum Erstellen, Aktualisieren und Löschen von in RDF repräsentierten Informationen abdeckt. Dieser Beitrag erläutert den Hintergrund und die Motivation der Entwicklung, erläutert technische Herausforderungen und gibt einen Überblick über den gegenwärtigen Stand des Projekts.

## Motivation und erwarteter Nutzen

Aus Sicht des hbz als Verbundzentrale stellt Linked Open Data einen sehr interessanten Ansatz für die zukünftige verteilte, kooperative Katalogisierung im Web dar. Einige Vorteile seien im Folgenden genannt.

### Webintegration

Die Berücksichtigung internationaler domänenübergreifender Web-Standards bei der Erstellung und Publikation von Daten bringt es mit sich, dass die resultierenden Daten in das Web integriert sind. Das bedeutet konkret:

- *Erhöhte Auffindbarkeit.* In das Web integrierte Daten können leicht von Suchmaschinen, Portalen wie Europeana oder Die Deutsche Bibliothek sowie anderen Aggregatoren eingesammelt werden.

---

<sup>1</sup> Zur Klarstellung unserer Verwendung des Ausdrucks “Linked Open Data”: Wir bezeichnen eine Datenpublikation als “Linked Data” wenn sie den Anforderungen in Berners-Lee, 2006 genüge trägt. Als “Linked Open Data” bezeichnen wir publizierte Daten, wenn sie zusätzlich zu den genannten Anforderungen Open Knowledge Foundation, 2009 entsprechen. Für mehr Informationen zu Open Data siehe Pohl, 2010. Linked Data wird in Pohl, 2011 detailliert erläutert.

- *Vereinfachte Mehrfachnutzung.* Publiizierte RDF-Daten können ohne großen Aufwand von verschiedenen Services parallel genutzt werden – inner- wie außerhalb des hbz.
- *Maximale Interoperabilität und Nachnutzbarkeit.* Webstandards erleichtern die Nachnutzung von Daten, indem sie notwendige Konversionsprozesse wie auch das Post-Processing minimieren.
- *Flexibilität.* RDF und Triple Stores sind sehr flexibel im Hinblick auf Erweiterungen und Änderungen des zugrundeliegenden Datenmodells.

#### Synergieeffekte

Die Anwendung von Best Practices aus der Linked-Open-Data-Community und der sukzessiven Standardisierung der von verschiedenen hbz-Services produzierten Daten bewirkt intra- wie interinstitutionelle Synergieeffekte. Bei der gegenseitigen Nachnutzung von Daten verschiedener hbz-Projekte sehen wir dies bereits jetzt. LOD macht dies auf unkomplizierte Weise möglich, wohingegen die Berücksichtigung proprietärer Schnittstellen und verschiedener Formate eine Kommunikation verschiedener Services recht ressourcenaufwändig macht. Somit hat eine Standardisierung von Services den Effekt, Ressourcen freizusetzen, die für die Optimierung bestehender Dienstleistungen oder den Aufbau zusätzlicher Angebote genutzt werden können.

#### Größere Herstellerunabhängigkeit

Die Bereitstellung vieler bibliothekarischer Dienstleistungen ist technologieabhängig und es gibt immer weniger Anbieter und damit auch weniger Konkurrenz in dem Geschäft mit bibliothekarischer Software. Viele Institutionen sind stark abhängig von den Produkten nur eines Anbieters und ein Wechsel zu einem anderen Anbieter ist mit signifikanten Kosten verbunden, mit anderen Worten: Es existiert ein starker Lock-In-Effekt.

Die Anwendung von Best Practices der Linked-Open-Data-Community verbessert diese Situation für die Kunden (Bibliotheken und Verbünde) auf zweierlei Weise:

1. Ist man in der Lage, die eigenen Daten leicht in einem offenen, domänenübergreifenden Standard aus einem System zu exportieren, erleichtert dies einen Umstieg auf ein anderes System erheblich, wodurch ein Lock-In-Effekt verhindert wird.
2. Werden bibliothekarische Daten in RDF und nicht mehr in – zwar offenen aber opaken – Formaten wie MARC oder MAB gespeichert und ausgetauscht, dann wird auch andere, generische Software von anderen Anbietern interessant

bzw. es wird für andere Anbieter leichter den Markt zu betreten. Ein stärkerer Wettbewerb führt zur Verbesserung der Produkte und/oder zu sinkenden Preisen.

### **lobid.org**

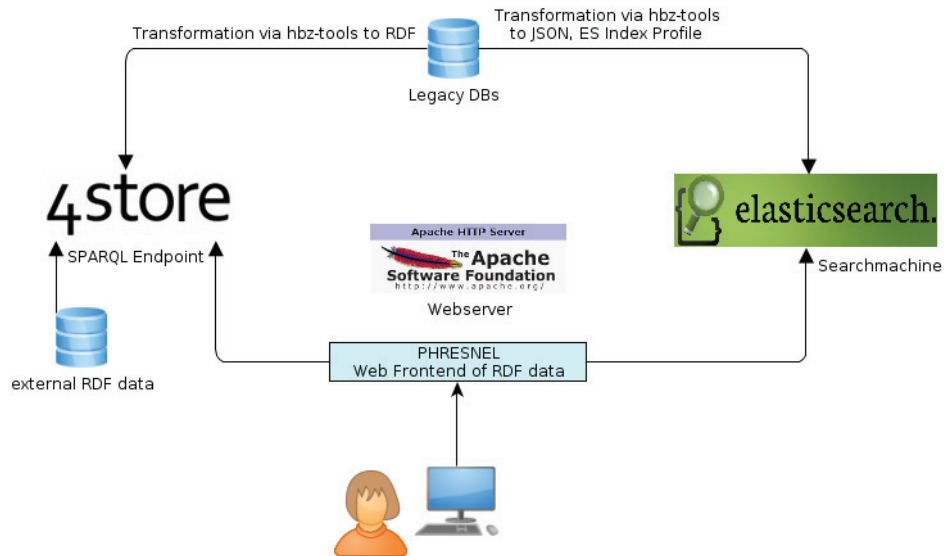
Im August 2010 startete das hbz den experimentellen Linked-(Open)-Data-Dienst lobid.org<sup>2</sup>, der aus zwei Subdiensten besteht: bibliographische Ressourcen (lobid-resources) und einem Verzeichnis von Bibliotheken, Museen und verwandten Organisationen (lobid-organisations). lobid.org folgt den Linked-Data-Prinzipien und geht auch – wann immer möglich – konform mit der Open Definition (Open Knowledge Foundation, 2009).

Seit 2010 wurden und werden die beiden lobid.org-Services und die ihnen zugrundeliegenden RDF-Daten kontinuierlich verbessert:

- Die Daten wurden durch mehrfache Überarbeitung von Konvertierung und Vokabularauswahl verbessert.
- Die vorhandene Information wurde zunehmend kontextualisiert durch Links zu anderen Datenquellen.
- Interaktionsmöglichkeiten für Endnutzer wurden verbessert, z.B. durch eine Suchmaschinenanbindung und die Vereinheitlichung der Oberfläche beider Subdienste (siehe Abschnitt 5).

---

<sup>2</sup><http://lobid.org/>



All marks mentioned may be trademarks or registered trademarks of their respective owners.

**Abbildung 1: lobid.org Technologie-Stack**

Abbildung 1 gibt einen allgemeinen Überblick über den gegenwärtig in lobid.org verwendeten Technologie-Stack, die Datenquellen und Konversionsprozesse. Es wird ausschließlich Open-Source-Software eingesetzt. Der verwendete Triple Store ist 4store<sup>3</sup> der Firma Garlik, für die Volltextindexierung und die Suchschnittstelle wird elasticsearch<sup>4</sup> eingesetzt. Das Webfrontend läuft auf einem Apache-Server und wird mit dem Phresnel-Framework generiert, das unten detailliert beschrieben wird.

Wie in Abbildung 1 zu sehen ist, basiert lobid.org fast ausschließlich auf Daten, die aus existierenden Legacy-Systemen exportiert und – unter Nutzung hbz-spezifischer Tools – nach RDF konvertiert werden. Die resultierenden RDF-Daten werden angereichert mit Links zu anderen Datenquellen in der LOD-Cloud. Darüber hinaus werden auch einige externe LOD-Datasets in den Triple Store geladen: dies sind derzeit die in lobid.org genutzten Ontologien sowie die von der deutschen

<sup>3</sup> <http://4store.org/>

<sup>4</sup> <http://www.elasticsearch.org/>

Nationalbibliothek (DNB) bereitgestellte Gemeinsame Normdatei (GND)<sup>5</sup>. Bisher gibt es keine Möglichkeit, auf einfache Weise manuell neue Informationen zu ergänzen oder Fehler in den Daten zu beheben.

### lobid organisations

Als das hbz mit der Publikation von Linked Open Data begann, wurde schnell klar, dass die bibliographischen Daten und Bestandsinformationen aus dem hbz-Verbundkatalog nur einen Anfang darstellen würden. Um nützliche Services auf der Basis von Linked Open Data aufzubauen, ist es ebenso notwendig URIs für und RDF-Beschreibungen von Exemplaren, Bibliotheken und Dienstleistungen zu haben. Zum Beispiel braucht man für eine geobasierte Suche nach Exemplaren einer bestimmten Ausgabe im Umkreis von 5 km URIs und Beschreibungen von mindestens drei Dingen: einer Manifestation M, die exemplifiziert wird durch ein Exemplar E, das im Bestand einer Organisation O ist. Als RDF-Graph sieht das beispielhaft wie folgt aus (Abbildung 2):

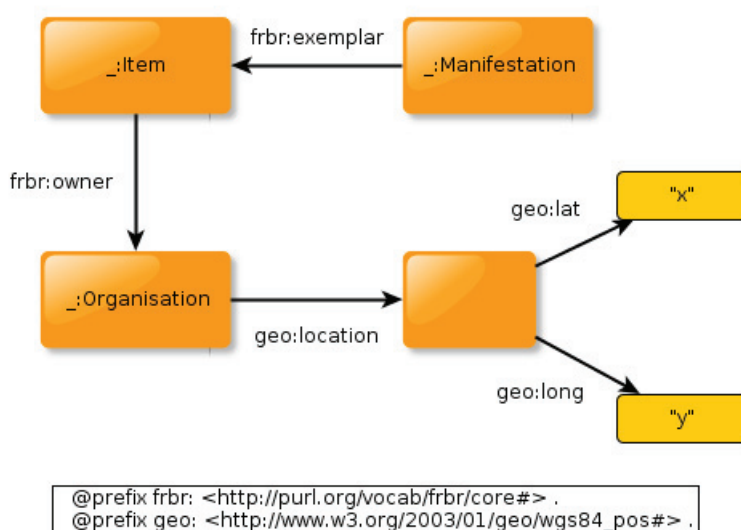


Abbildung 2: Graph der Relationen zwischen Manifestation, Exemplar und Organisation

---

<sup>5</sup> Siehe den Eintrag für das LOD-GND-Dataset auf the Data Hub: <http://thedatahub.org/dataset/dnb-gemeinsame-normdatei>.

Da sich vor zwei Jahren wenig bewegte, wenn man Menschen und/oder Organisationen darum bat, Linked Data bereitzustellen, wurde kurzerhand beschlossen, selbst URIs für bibliothekarische Institutionen zu prägen und Beschreibungen bereitzustellen. Im Juli 2010 startete schließlich lobid.org mit dem Subdienst lobid-organisations (Ostrowski, 2010).

Die dem Organisationsverzeichnis zugrundeliegenden Datenquellen sind das deutsche ISI- und Sigelverzeichnis<sup>6</sup> und die MARC Organisation Code Database<sup>7</sup>. Bislang wurden in lobid.org URIs für mehr als 40.000 Institutionen geprägt und mit entsprechenden Informationen<sup>8</sup> versehen. Allerdings wird weder ein Vollabzug der Daten zum Download bereitgestellt noch sind die Daten offen lizenziert, weil die Daten nicht vom hbz produziert wurden und es somit nicht berechtigt ist eine Nutzungslizenz für Dritte zu bestimmen.

Zusätzlich zu den Daten aus den erwähnten Datenquellen werden Verlinkungen zu anderen Datenquellen in der LOD-Cloud generiert. Bis jetzt wurden Links zur DBpedia und Wikipedia (Christoph, 2012d) sowie zu GeoNames (Christoph, 2012a) ergänzt. Darüberhinaus wurden die Organisationsbeschreibungen mit QR-Codes angereichert, in denen Kontaktinformationen kodiert sind (ebd.). Außerdem existieren Geokoordinaten für die meisten Institutionen, so dass etwa ein entsprechender Kartenausschnitt in lobid.org eingebettet werden kann.<sup>9</sup> Eine Beispielbeschreibung (die - wie in Abschnitt 5.1 beschrieben - auf Basis der RDF-Daten generiert wurde) wird in Abbildung 3 gezeigt.

---

<sup>6</sup><http://dispatch.opac.d-nb.de/DB=1.1/DB=1.2/<>

<sup>7</sup><http://loc.gov/marc/organizations/>

<sup>8</sup>Für einen Überblick über die in lobid-organisations benutzten RDF-Properties siehe <http://lobid.org/dataset/organisation/properties.txt>.

<sup>9</sup>Diese Kartenausschnitte basieren auf den Daten und dem Service von <http://www.openstreetmap.org/>.



<a href="http://lobid.org/organisation/DE-101">http://lobid.org/organisation/DE-101</a>									
Name	Deutsche Nationalbibliothek								
Geokoordinaten									
Anschrift	<table border="1"> <tr> <td>Straße</td><td>Adickesallee 1</td></tr> <tr> <td>PLZ</td><td>60322</td></tr> <tr> <td>Ort</td><td>Frankfurt/Main</td></tr> <tr> <td>Land</td><td>Germany</td></tr> </table>	Straße	Adickesallee 1	PLZ	60322	Ort	Frankfurt/Main	Land	Germany
Straße	Adickesallee 1								
PLZ	60322								
Ort	Frankfurt/Main								
Land	Germany								
ISIL	DE-101								
Homepage	<a href="http://www.dnb.de">http://www.dnb.de</a>								
E-Mail	<a href="mailto:info-f@dnb.de">mailto:info-f@dnb.de</a>								
Kontakt-QR									
Weitere Information	<a href="http://dispatch.opac.d-nb.de/DB=1.2/CMD?ACT=SRCHA&amp;IKT=8529&amp;TRM=DE-101">http://dispatch.opac.d-nb.de/DB=1.2/CMD?ACT=SRCHA&amp;IKT=8529&amp;TRM=DE-101</a>								
Sitz	<a href="http://sws.geonames.org/6553153/">http://sws.geonames.org/6553153/</a>								
Wikipedia-Link	<a href="http://de.wikipedia.org/wiki/Deutsche_Nationalbibliothek">http://de.wikipedia.org/wiki/Deutsche_Nationalbibliothek</a>								
DBpedia-Link	<a href="http://de.dbpedia.org/resource/Deutsche_Nationalbibliothek">http://de.dbpedia.org/resource/Deutsche_Nationalbibliothek</a>								

Abbildung 3: lobid.org-Beschreibung der Deutschen Nationalbibliothek<sup>10</sup>

### lobid resources

lobid-resources stellt eine Linked-Open-Data-Schnittstelle für Open Data aus der hbz-Verbunddatenbank dar. Publiziert werden URIs für und Beschreibungen von bibliographischen Ressourcen wie Monographien, Zeitschriften, Serien, Mehrbändigen Werken usw. Seit der ersten Publikation von Open Data im hbz-Verbund im März 2010 (Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, 2010) wurden immer mehr Daten des Verbundkatalogs unter einer offen Lizenz veröffentlicht. Mit Stand August 2012 kooperieren bis auf zwei Bibliotheken alle hbz-Verbundbibliotheken mit dem hbz bei der Freigabe der Daten (Christoph,

<sup>10</sup> URL: <http://lobid.org/organisation/DE-101/about>

2012). Damit umfassen die Daten mittlerweile knapp 16 Millionen Titeldatensätze, die unter einer Creative-Commons-Zero-Lizenz<sup>11</sup> veröffentlicht werden, das sind etwa 85% des gesamten Verbundkatalogs.<sup>12</sup> Die Daten können über eine Suchschnittstelle<sup>13</sup> sowie einen SPARQL-Endpoint<sup>14</sup> abgefragt werden. Außerdem steht ein Vollabzug der RDF-Daten zum Download<sup>15</sup> bereit.

Auch die lobid-resources Daten wurden mit Verlinkungen zu anderen Datenquellen angereichert. Mittels einfacher ISBN- und Titel-Matchingalgorithmen und ein wenig Postprozessierung wurden für einige Ressourcen Links zur Dbpedia (Christoph, 2012b), zur Open Library (Christoph, 2012c) und zu Projekt Gutenberg ergänzt. Diese Links sind über die einfache Verlinkung hinausgehend nützlich. Beispielsweise wird eine wechselseitige Anreicherung verschiedener Titel mit Schlagwörtern, Links etc. ermöglicht.

Abbildung 4 zeigt die Beispielbeschreibung einer bibliographischen Ressource in lobid.org inklusive Links zu anderen Datenquellen.

---

<sup>11</sup> <http://creativecommons.org/publicdomain/zero/1.0/>

<sup>12</sup> Neben Daten der Bestände der zwei Bibliotheken, die nicht mit dem hbz kooperieren, werden auch die Metadaten, zu den über die Nationallizenzen erworbenen Inhalten nicht als Open Data publiziert. Der Grund ist, dass die von der DFG festgelegten "Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen", denen die einzelnen mit den Verlagen im Rahmen der Nationallizenzen geschlossenen Verträgen entsprechen müssen, eine Publikation der Metadaten als Open Data verhindern (vgl. Kreutzer, 2011, S. 22f.).

<sup>13</sup> <http://lobid.org/resource/search>

<sup>14</sup> <http://lobid.org/sparql/>

<sup>15</sup> Siehe die Beschreibung im Open-Data-Verzeichnis the Data Hub: <http://thedatahub.org/dataset/lobid-resources>.



<a href="http://lobid.org/resource/HT002189125">http://lobid.org/resource/HT002189125</a>	
Titel	Typee
Titelzusatz	a peep at Polynesian life
Autor	<a href="http://d-nb.info/gnd/118580604">http://d-nb.info/gnd/118580604</a>
Erscheinungsjahr	1982
Typ	< <a href="http://purl.org/ontology/bibo/Book">http://purl.org/ontology/bibo/Book</a> >
Typ	< <a href="http://purl.org/vocab/frbr/core#Manifestation">http://purl.org/vocab/frbr/core#Manifestation</a> >
Volltext	<a href="http://gutenberg.org/ebooks/1900">http://gutenberg.org/ebooks/1900</a>
Volltext	<a href="http://gutenberg.org/ebooks/28656">http://gutenberg.org/ebooks/28656</a>
Volltext	<a href="http://gutenberg.org/ebooks/9269">http://gutenberg.org/ebooks/9269</a>
Volltext	<a href="http://gutenberg.org/ebooks/23969">http://gutenberg.org/ebooks/23969</a>
Ausgabe	3. print.
Erscheinungsort	New York, NY
Verlag	Literary Classics of the United States [u.a.]
Format	print
Umfang	1333 S.
ISBN-10	0521262194
ISBN-10	0940450003
ISBN-13	9780521262194
ISBN-13	9780940450004
Werktitel	<a href="http://openlibrary.org/works/OL14953734W">http://openlibrary.org/works/OL14953734W</a>
Werktitel	<a href="http://dbpedia.org/resource/Typee">http://dbpedia.org/resource/Typee</a>
Werktitel	<a href="http://de.dbpedia.org/resource/Typee">http://de.dbpedia.org/resource/Typee</a>
Exemplar	<a href="http://lobid.org/item/HT002189125%3AU+MEL-11">http://lobid.org/item/HT002189125%3AU+MEL-11</a>
	Typ < <a href="http://purl.org/vocab/frbr/core#Item">http://purl.org/vocab/frbr/core#Item</a> >
	Zugehöriger Titel <a href="http://lobid.org/resource/HT002189125">http://lobid.org/resource/HT002189125</a>
	Besitzer <a href="#">Institut für Anglistik, Amerikanistik und Keltologie, Bibliothek</a>
	Signatur U MEL-11
Weitere Information	<a href="http://193.30.112.134/F/?func=find-c&amp;ccl term=IDN%3DHT002189125">http://193.30.112.134/F/?func=find-c&amp;ccl term=IDN%3DHT002189125</a>

Abbildung 4: Beispielbeschreibung einer bibliographischen Ressource in lobid.org<sup>16</sup>

<sup>16</sup> URL: <http://lobid.org/resource/HT002189125/about>

## Fresnel Display Vocabulary for RDF

### Hintergrund

Anfangs wurden in lobid.org die konvertierten Daten anhand von Pubby, einem Linked Data Frontend für SPARQL Endpoints<sup>17</sup>, präsentiert. Wenngleich die Einrichtung von Pubby sich sehr einfach gestaltet, so exponieren die resultierenden Ansichten - insbesondere die für Menschen gedachte HTML-Darstellung - zu viele Details der zugrundeliegenden Technologien. Die Präsentation der Organisationsdaten in lobid.org hingegen wurde anhand von selbsterstellten SPARQL-Anfragen, PHP-Skripten und HTML-Templates umgesetzt. Diese Variante verhalf zwar zu größerer Flexibilität bei der Darstellung, allerdings war sie nicht ohne weiteres auf andere Daten übertragbar, da dafür jeweils eigene SPARQL-Anfragen und HTML-Templates benötigt wurden. Darüber hinaus stand bereits der Plan im Raum, Bibliotheken eine einfache Möglichkeit zu bieten, Informationen in RDFa zu erstellen. Um dies zu ermöglichen, wurde ein einfacher, intuitiver Editor benötigt. Anstelle direkt die zugrunde liegenden Daten zu bearbeiten, sollte also eine HTML-Formular-basierte Lösung eingesetzt werden, um so die Arbeit in einer vertrauten Web-Umgebung zu ermöglichen.

Mit diesen Anforderungen im Hinterkopf begann die Suche nach einer Schema-Sprache, anhand derer ein Datenmodell formuliert werden kann, aus welchem wiederum das Frontend automatisch hervorgeht. Beim Umgang mit RDF-Daten sind die naheliegendsten Lösungen RDF-Schema (RDFS) und die Web Ontology Language (OWL). Da aber die Ontologien, die in diesen Sprachen erstellt werden, meist anwendungsunabhängig sind, waren Experimente in diese Richtung eher unfruchtbar, da auch die resultierenden Editor-Ansichten zu generisch waren, um die Anforderungen zu erfüllen. Insbesondere die Verwendung von Klassen und Eigenschaften aus verschiedenen Vokabularen ist mit RDFS bzw. OWL nahezu unmöglich auf eine präzise und eingängige Art und Weise zu gestalten. Anders ist dies im *Fresnel Display Vocabulary for RDF*. Anhand dieses Vokabulares ist es möglich zu spezifizieren, *welche* Informationen aus einem RDF-Graphen angezeigt,

---

<sup>17</sup> <http://www4.wiwiiss.fu-berlin.de/pubby/>

und *wie* diese dargestellt werden sollen<sup>18</sup>, und zwar ohne mit den zugrunde liegenden Ontologien zu interferieren. Ähnlich wie auch die oben erwähnten Ontologie-Sprachen basiert auch Fresnel auf RDF. Dies ermöglicht es, innerhalb der gesamten Implementation in einem Datenmodel zu verbleiben.

#### Fresnel-Linsen

Fresnel-Linsen dienen der Spezifizierung, *welche* Daten angezeigt werden sollen. Eine einzelne Linse kann auf verschiedene Weisen mit Instanzen verknüpft werden. Die einfachste Möglichkeit ist dabei die Referenzierung der Klasse, also über die `rdf:type`-Eigenschaft, wie unten veranschaulicht. Für die so ausgewählten Instanzen wird dann eine geordnete Liste von Prädikaten angegeben, was dank der Turtle-Notation sehr einfach und gut lesbar ist. Um auch Daten zu einer verknüpften Entität mit anzuzeigen, kann anstelle eines einfachen Prädikates auch auf eine andere Linse verwiesen werden. Im Beispiel (Abbildung 5) etwa werden Vor- und Nachname des Autors ebenfalls angezeigt, und nicht nur dessen URI. Letzteres wäre der Fall, wenn nur das Prädikat `dc:creator` in der Liste aufgeführt würde, ohne Verweis auf die entsprechende `:person`-Sub-Linse wie im Beispiel auf der rechten Seite.

<pre>:person   a fresnel:Lens ; fresnel:classLensDomain foaf:Person ; fresnel:showProperties (   foaf:familyname   foaf:givenname ) .</pre>	<pre>:document   a fresnel:Lens ; fresnel:classLensDomain bibo:Document ; fresnel:showProperties (   dc:title   bibo:isbn   [ a fresnel:propertyDescription ;     fresnel:property dc:creator ;     fresnel:sublens :person ] ) .</pre>
---	---

Abbildung 5: Fresnel-Linsen und Sub-Linsen

---

<sup>18</sup> <http://www.w3.org/2005/04/fresnel-info/>

Die Anwendung obiger Linse auf einen Triplestore, der die entsprechenden Daten enthält, sollte die in Abbildung 7 dargestellten Tripel hervorbringen, welche genau in dieser Reihenfolge dargestellt werden sollen:

```
<http://lobid.org/resource/HT002948556>
  a bibo:Document ;
  dc:title "With reference to reference"@en ;
  bibo:isbn "9780915145539" ;
  dc:creator <http://d-nb.info/gnd/135539897> .

<http://d-nb.info/gnd/135539897>
  a foaf:Person ;
  foaf:familyName "Elgin" ;
  foaf:givenName "Catherine" .
```

**Abbildung 6: Auf Basis einer Fresnel-Linse gewonnene RDF-Daten**

Zusammenfassend lässt sich festhalten, dass Fresnel-Linsen eine sehr präzise und deklarative Möglichkeit bieten, um auszudrücken, welche Daten ausgewählt und in welcher Reihenfolge sie angezeigt werden sollen.

### Formate

Fresnel-Formate erlauben es auszudrücken, wie die anhand von Linsen ausgewählten Daten dargestellt werden sollen. Die Möglichkeiten erstrecken sich dabei von benutzerdefinierten Bezeichnungen von Prädikaten bis hin zu Referenzen auf externe CSS-Klassen. Ein Beispiel findet sich in Abbildung 7.

```
[] a fresnel:Format ;
  fresnel:propertyFormatDomain foaf:familyName ;
  fresnel:label "Last name"@en ;
  fresnel:label "Nachname"@de .

[] a fresnel:Format ;
  fresnel:classFormatDomain foaf:Person ;
  fresnel:resourceStyle "withBorder"^^fresnel:styleClass .
```

**Abbildung 7: Formate für benutzerdefinierte Prädikat-Bezeichnungen und anzuwendende CSS-Klassen**

Ähnlich wie es Fresnel-Linsen erlauben, Daten auszuwählen und zu ordnen, bieten Fresnel-Formate also die Möglichkeit, die Darstellung auf eine deklarative, anwendungsunabhängige Art und Weise zu konfigurieren.

### Implementation

Unter den verschiedenen existierenden Implementierungen<sup>19</sup> von Fresnel gab es keine mit einer Bearbeitungsansicht und keine die – wie lobid.org – in PHP implementiert ist. Außerdem schienen die meisten Projekte inaktiv. Daher wurde prototypisch ein neues PHP-basiertes Framework names “Phresnel” implementiert, das auch eine Datenbearbeitung ermöglicht. Derzeit wird nur eine kleine Untermenge der Linsen- und Formatfeatures unterstützt, der Umfang beschränkt sich hierbei auf die für einen Prototypen absolut notwendigen Funktionen.

### Anzeige

Um Daten entsprechend einer Fresnel-Linse anzuzeigen, ermittelt die Web-Anwendung zunächst anhand der URL die zu verwendende Linse, z.B "Dokument". Anschließend wird mithilfe des Phresnel-Frameworks eine generische (X)HTML-Ansicht (mit eingebettetem RDFa) der angeforderten Daten erzeugt (Abbildung 8). Zu diesem Zeitpunkt wird davon ausgegangen, dass ein HTTP-303-Redirect entsprechend des Linked-Data-Design-Patterns<sup>20</sup> bereits stattgefunden hat.

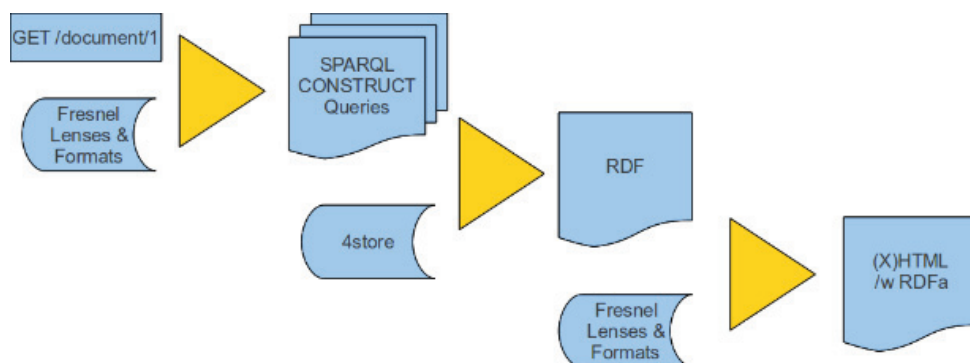


Abbildung 8: Generierung von (X)HTML-Ansichten mit Phresnel

---

<sup>19</sup> <http://www.w3.org/2005/04/fresnel-info/#implementation>

<sup>20</sup> <http://www.w3.org/TR/cooluris/#r303gendocument>

Intern verwendet Phresnel die Linsen-Definitionen, um wie in Abbildung 9 dargestellt eine Reihe von SPARQL-CONSTRUCT-Abfragen zu generieren. Danach wird entsprechend der Linsen- und Format-Definitionen die eigentliche Ausgabe erzeugt.

```
CONSTRUCT {
  <http://lobid.org/resource/HT002948556> ?p ?o .
} WHERE {
  <http://lobid.org/resource/HT002948556> a bibo:Document .
  <http://lobid.org/resource/HT002948556> ?p ?o .
  FILTER (?p=rdf:type || ?p=dc:title || ?p=bibo:isbn || ?p=dc:creator) .
}

CONSTRUCT {
  ?s ?p ?o .
} WHERE {
  <http://lobid.org/resource/HT002948556> dc:creator ?s .
  ?s a foaf:Person .
  ?s ?p ?o .
  FILTER (?p=rdf:type || ?p=foaf:givenName || ?p=foaf:familyName) .
}
```

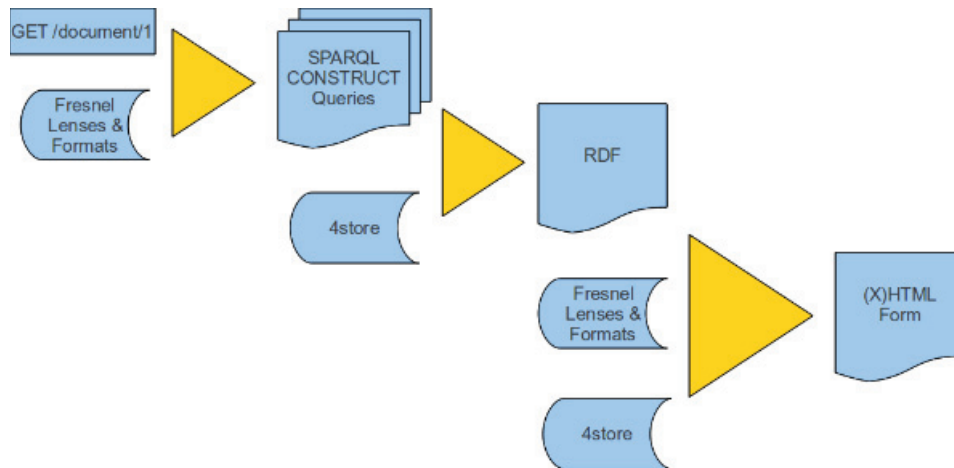
**Abbildung 9: Von Phresnel generierte SPARQL-CONSTRUCT-Abfragen**

Derzeit ist nur ein hartcodiertes Box-Model (Bizer et al., 2005) auf Basis von verschachtelten Tabellen implementiert. Die Transformation zu (X)HTML kann natürlich übersprungen werden, wenn anhand von Content Negotiation eine reine RDF-Darstellung angefragt wurde.

### Bearbeiten

Die Schritte für die Erstellung der Bearbeitungsansicht einer Ressource sind den obigen, zur reinen Darstellung notwendigen, sehr ähnlich – mit einem großen Unterschied. Prinzipiell wird das RDF, welches sich durch die Auswertung der erzeugten SPARQL-Abfragen ergibt, zu einem (X)HTML-Formular umgewandelt, indem alle Literale einfach durch Texteingabeelemente ersetzt werden. Dies führt allerdings in den Fällen, in denen Links zu anderen Entitäten erzeugt werden sollen,

zu einem unschönen Interface, da hier URIs eingegeben werden müssten. Mit Hinsicht auf die Benutzbarkeit ist dies inakzeptabel.



**Abbildung 10: Generierung von (X)HTML-Formularen mit Phresnel**

Die aktuelle Implementierung löst dieses Problem auf eine nur ausreichende, aber zumindest einigermaßen optisch ansprechende Art und Weise. Verfügbare Link-Ziele aus dem Triple Store, z. B. dem obigen Beispiel folgend eine Liste der möglichen Autoren, werden als Dropdown-Listen angezeigt. Leider bringt diese Lösung zwei schwerwiegende Einschränkungen mit sich: Die Liste kann (1) zu lang werden und (2) sind die Listeneinträge auf die eigene Datenquelle beschränkt, was es schwierig macht, andere Quellen im Web zu verknüpfen. Für die Zukunft ist geplant, diese beiden Einschränkungen mit Hilfe von Autovervollständigung zu beheben. Diese wird anhand von Javascript gesteuert und soll es ermöglichen, sowohl für interne als auch externe Datenquellen die Eingabe des Nutzers zu vervollständigen, so dass eine komplette Liste der möglichen Verknüpfungen nicht mehr notwendig ist. Für das Verlinken auf interne Daten kann hierbei der im nächsten Abschnitt erläuterte Suchindex verwendet werden. An dieser Stelle sei nochmals darauf hingewiesen, dass zwar Daten mit diesem Web-Frontend bearbeitet werden können, zur Zeit aber kein Workflow für die Speicherung implementiert ist.

## Suche

Während das Linked-Data-Paradigma für die Navigation hervorragend geeignet ist, so sind die Möglichkeiten einer Volltextsuche für das Auffinden von Ressourcen unabdingbar. Zwar unterstützt SPARQL reguläre Ausdrücke<sup>21</sup>, die für die Volltextsuche verwendet werden können. Dies skaliert allerdings derzeit nicht auf die Datenmenge von lobid.org.

Deshalb basiert sowohl die Suche für Organisationen<sup>22</sup> als auch die für Ressourcen<sup>23</sup> derzeit auf einem elasticsearch-Index<sup>24</sup>, der wiederum über eine API, die eine CQL-Schnittstelle bereitstellt, angesprochen wird. Dieses konkrete Setup hat historische Gründe.<sup>25</sup>

Wegen dieses Aufbaus haben die aus der CQL-Anfrage resultierenden Suchergebnisse, die als Atom-Feeds ausgeliefert werden, eine von den Linsen-Definition abweichende Datenstruktur. Daher werden derzeit nur Identifikatoren aus den Suchergebnissen extrahiert, um die entsprechenden Ressourcen dann anhand von Phresnel aus dem Triplestore zu laden und darzustellen. Dies funktioniert recht gut und ist deutlich leistungsfähiger als die Verwendung von nativen SPARQL-Abfragen für die Volltextsuche. Die Tatsache, dass, nachdem die Ressource identifiziert worden ist, die entsprechenden Daten noch aus dem Triple Store geholt werden müssen, zeigt, dass dieser Aufbau aber noch nicht optimal ist. Besser wäre es, die Daten direkt aus dem Suchergebnis zu übernehmen, was aber eben eine Harmonisierung der Datenstrukturen zwischen Volltextindex und Triplestore voraussetzt. Dies ist nur eines der Probleme, die als nächstes in Angriff genommen werden sollen.

---

<sup>21</sup> <http://www.w3.org/TR/rdf-sparql-query/#funcex-regex>

<sup>22</sup> <http://lobid.org/organisation/search>

<sup>23</sup> <http://lobid.org/resource/search>

<sup>24</sup> <http://www.elasticsearch.org/>

<sup>25</sup> Der Titeldaten-Index hat lange vor lobid.org existiert und wird von mehreren hbz-Diensten verwendet. Es war am einfachsten, die Index für Organisationen einfach in die gleiche Infrastruktur zu integrieren.



## **Ausblick**

Die derzeitige Proof-of-Concept-Implementierung eines LOD-basierten System für den Lese- und Schreibzugriff auf Bibliotheksdaten zeigt interessante Perspektiven für das zukünftige Datenmanagement auf. Unbestreitbar sind die Bemühungen noch ganz am Anfang. Möglichkeiten zur Weiterentwicklung des Systems werden im folgenden aufgezeigt.

### Rein Javascript-basierter Editor

Da der wichtigste Teil des Editors - die Komponente zum Auffinden und Verknüpfen von Ressourcen - aus oben genannten Gründen überarbeitet werden muss, wird in Betracht gezogen, den Editor insgesamt auf Basis von RDFa und Javascript zu implementieren. In einer create.js<sup>26</sup> ähnlichen Art würde damit auf serverseitig generierte Formulare zugunsten eines clientseitigen Editors verzichtet. Dies würde nicht nur die Benutzerfreundlichkeit verbessern, sondern auch die Belastung des Servers minimieren und für eine bessere Trennung zwischen Front- und Backend sorgen.

### Datenproduktion und Pflege

Die Ergebnisse einer Bearbeitung von Daten anhand des Web-Frontends müssen persistent gemacht werden. Es gibt einige nicht-triviale Entscheidungen, die in dieser Hinsicht getroffen werden müssen. So gilt es etwa zu klären, wo die Daten gespeichert werden (Triple Store, Dateisystem, Suchindex), wie sie organisiert werden (Named Graphs (Dodds, 2009), ...), welche Provenienzdaten benötigt werden und wie das Berechtigungssystem aufgebaut sein muss. Da eine Anwendung wie die in diesem Artikel beschriebene in einer vernetzten, dezentralen Umgebung verwendet wird, muss auch eine Lösung dafür angedacht werden, wie andere angeschlossene Dienste über die Erstellung, Aktualisierung und Löschung von Daten informiert werden können<sup>27</sup>. Eine Idee in dieser Hinsicht ist es, ein nachrichtenbasiertes Protokoll wie IRC oder XMPP<sup>28</sup> zu verwenden.

---

<sup>26</sup> <http://createjs.org/>

<sup>27</sup> Eine Initiative zur Entwicklung eines Synchronisierungsstandards für Webinhalte mit dem Namen "ResourceSync" gibt es bereits bei der NISO, siehe <http://www.niso.org/workrooms/resourcesync/>. Ein daraus resultierender Standard wäre sehr relevant für diesen Kontext.

<sup>28</sup> <http://xmpp.org/>

Offensichtlich wichtige Aspekte der Datenverwaltung sind die Identifikation und Authentifizierung derjenigen Akteure, die die Daten bearbeiten. Anstatt ein solches System von Grund auf neu zu konzipieren und implementieren, sollte möglichst auf bestehende Standards zurückgegriffen werden – im Idealfall sollten diese ebenfalls auf Linked-Data-Prinzipien fußen. Aus diesem Grund ist der vielversprechendste Ansatz der Einsatz von WebID (Sporny et al., 2011). Hierbei werden FOAF-Beschreibungen in Verbindung mit SSL-Zertifikaten verwendet, was zu einem sicheren verteilten Identifizierungs- und Authentifizierungsmechanismus führt, der sowohl für Menschen als auch für Maschinen nachvollziehbar ist.

#### Versionierung

Zu den wichtigsten Provenienzzangaben gehört ein lückenloser Verlauf der an den Daten vollzogenen Änderungen zusammen mit der Identifikation des Akteurs (sei es ein System oder eine Person), der für diese Änderungen verantwortlich ist. Während es durchaus möglich ist, Änderungen an RDF-Daten als Changesets anhand eines RDF-Vokabulars<sup>29</sup> auszudrücken, ist dies sehr wahrscheinlich nicht der effizienteste Weg, da für eine Änderung sehr viele Tripel benötigt werden. Zu den Alternativen, die in Erwägung gezogen werden, gehören der Einsatz eines Systems zur Versionsverwaltung das auf einfachen Dateien basiert (z.B. git<sup>30</sup>) und die Verwendung der Versionskontrollfeatures von elasticsearch. Um die verschiedenen Versionen der Daten in einer standardisierten Art und Weise ausliefern zu können ist darüber hinaus angedacht, eine Memento-Schnittstelle<sup>31</sup> für das ausgewählte Versionierungssystem zu implementieren.

#### JSON-LD in ES / Fresnel-basierter Suchindex

Die Art und Weise, auf die die Suche derzeit in die Anwendung eingebunden ist, ist nicht sehr generisch und hängt von externen organisatorischen und technischen Prozessen ab. Da aber der Lookup ein sehr wichtiger Teil des Gesamtsystems ist, wird eine besser eingebundene Lösung angestrebt. Elasticsearch als schemaloser Suchindex sollte mit RDF-Daten gut verträglich sein. Da es Daten auf Basis von

---

<sup>29</sup> Siehe etwa <http://docs.api.talis.com/getting-started/changesets>.

<sup>30</sup> <http://git-scm.com/>

<sup>31</sup> <http://www.mementoweb.org/>

JSON indiziert, werden verschiedene JSON-RDF-Serialisierungen evaluiert. Der vielversprechendste Ansatz scheint momentan JSON-LD<sup>32</sup> sein. Zum einen wird diese Serialisierung derzeit zum Standard für die Codierung von RDF in JSON, zum anderen – und dies ist aus technischer Sicht wesentlich wichtiger – verwendet JSON-LD einen Key-Value-Ansatz, der mit elasticsearch sehr gut zusammenspielt. Die bereits für die Darstellung und Bearbeitung von Ressourcen verwendeten Fresnel-Linsen könnten verwendet werden, um die JSON-LD-Strukturen für den Index zu erzeugen.

### Literaturverzeichnis

*[Für alle URLs gilt: Letzter Zugriff am 30.6.2012.]*

T. Berners-Lee (2006, 27. Juli). Linked Data - Design Issues. URL: <http://www.w3.org/DesignIssues/LinkedData.html>

C. Bizer, R. Lee & E. Pietriga (2005, 30. Juni). Fresnel - Display Vocabulary for RDF [User Manual]. URL: <http://www.w3.org/2005/04/fresnel-info/manual/>

P. Christoph (2012a, 16. April). lobid.org news [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/mYQ7>

P. Christoph (2012b, 3. Mai). First results using SILK to link to Dbpedia [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/mAJH>

P. Christoph (2012c, 23. Mai). 1.2 M links to Open Library [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/l4BO>

P. Christoph (2012d, 4. Juni). lobid-organisation - new links to German Dbpedia [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/6lBO>

P. Christoph (2012e, 16. August). lobid-resources - much more records, authority URIs dereferencable, provenance [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/cQJ5>

Conference of European National Librarians (2011, 28. September). Europe's national librarians support Open Data licensing [Pressemitteilung]. URL: <https://app.e2ma.net/app/view:CampaignPublic/id:1403149.7214447972/rid:48e64615892ac6adde9a4066e88c736c>

---

<sup>32</sup> <http://json-ld.org/>

L. Dodds (2009, 5. November). Managing RDF Using Named Graphs [Blogbeitrag]. URL: <http://www.ldodds.com/blog/2009/11/managing-rdf-using-named-graphs/>

D. Marcum & Library of Congress (2011, 31. Oktober). A Bibliographic Framework for the Digital Age. URL: <http://www.loc.gov/marc/transition/news/framework-103111.html>

Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010). Köln, im März 2010: Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit [Pressemitteilung]. URL: <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>

T. Kreutzer (2011). Open Data – Freigabe von Daten aus Bibliothekskatalogen. Ein Leitfaden. Hg. v. Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen. URL: <http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>

Open Knowledge Foundation (2009, November). Open Knowledge Definition, v1.1. URL: <http://opendefinition.org/okd/>

F. Ostrowski (2010, 20. Juli). Building a Linked Data based index of library institutions [Blogbeitrag]. URL: <http://blog.lobid.org/2010/07/building-linked-data-based-index-of.html>

A. Pohl (2010): Open Data im hbz-Verbund. ProLibris, 3, S. 109-113. Preprint u.a. einsehbar unter <http://hdl.handle.net/10760/14838>

A. Pohl (2011): Linked Data und die Bibliothekswelt. Erscheint im Konferenzband zur ODOK 2010. Preprint einsehbar unter <http://hdl.handle.net/10760/15324>.

M. Sporny, T. Inkster, H. Story, B. Harbulot & R. Bachmann-Gmür (2011, 12. Dezember). WebID 1.0 - Web Identification and Discovery [W3C Editor's Draft]. URL: <http://www.w3.org/2005/Incubator/webid/spec/>

---

# **Vernetztes Wissen – Menschen**



## Das Netzwerk und die Institution

Lambert Heller

TIB/UB Hannover

*Anmerkungen zum Trend der "Vernetzung" von Aufmerksamkeit und Engagement in der Bibliotheks- und Informationsbranche*

### Zusammenfassung

Mit Barry Wellman et al. wird zunächst beschrieben, wie der Aufstieg des "Vernetzten Individualismus" durch Handy, Internet und Facebook das soziale Leben verändert hat. Clay Shirky folgend birgt dies ein Potential neuartiger Organisation und Artikulation vernetzter Individuen. In Gestalt heterogener Projekte, die informell von vernetzten Einzelnen (statt von etablierten Institutionen) initiiert werden, taucht dieses Phänomen nun auch in der BID-Branche auf. In Deutschland herrscht hier jedoch Nachholbedarf - auch, weil den dazugehörigen neuen Medien in Ausbildung und Beruf bisher oftmals die Anerkennung versagt bleibt. Neugier, Lernbereitschaft und Offenheit für Eigeninitiative innerhalb der Institution (Intrapreneurship, hier nach Adam Sofronijevic), so der abschließende Appell, sind notwendig, um das vernetzte Engagement und Know How in die Institutionen einzubringen und dort zu kultivieren.

### Abstract

Following Barry Wellman et al., the rise of "networked individualism" is described, as a change of our social lives through the dissemination of cell phones, Internet and Facebook. Clay Shirky claims the potential of a new kind of individualist, networked organizing and articulation rising from this kind of broad media evolution. In the libraries, archives and documentation sector, we already see a rise of heterogeneous projects, initiated by networked individuals (rather than traditional organizations). In Germany, organizations still have to catch up with this development; with one reason for that maybe the lack of recognition of networked media by colleges and employers. Curiosity, willingness to learn and openness towards self-initiative inside the workplace / organization (Intrapreneurship, according to Adam Sofronijevic) will be necessary to integrate and cultivate the networked engagement and know how on the inside of established organizations.



*1. Mehr als nur Individualisierung und Mobilität: Wie der Aufstieg des "Vernetzten Individualismus" durch Handy, Internet und Facebook unser soziales Leben verändert hat. "Weak Ties" und das Empowerment der (bisherigen) Informationskonsumenten.*

Der kanadische Soziologe Barry Wellman und seine KollegInnen beschreiben, teils basierend auf Daten des Pew Research Centers, den "vernetzten Individualismus" als einen raschen Bedeutungsverfall einstmals stabiler sozialer Kontexte wie Familie, Dorf oder auch urbaner Nachbarschaften, in denen das Individuum einstmals primär eingebunden war, und die seinen Zugriff auf die Welt und auf andere Menschen bestimmten. Wellman beschreibt eine allgemeine Zunahme von Individualität und Mobilität, die in den letzten zwanzig Jahren das Aufwachsen und den Alltag vieler Menschen radikal verändert haben. (Wellman u. a. 2003)

So ist für EuropäerInnen mit Zugang zu akademischer Bildung das Auslandssemester zum Normalfall geworden, und für praktisch die gesamte Bevölkerung des globalen Nordens haben Billigflieger weite Reisen normal gemacht. Von unumkehrbarer Bedeutung für den vernetzten Individualismus sind demzufolge jedoch vor allem drei technische Entwicklungen: Die Popularisierung des Internets und des Handys seit den Neunziger Jahren, sowie in den darauffolgenden Jahren das Aufkommen sozialer Werkzeuge wie Facebook.

Deutlich wird dies Wellman zufolge schon am Telefon. Erreichte man bis vor zwanzig Jahren unter einer Telefonnummer typischerweise einen Haushalt, dessen Mitglieder sich den Telefonanschluss teilen, ist der Normalfall heute, dass jeder jederzeit und einer individuellen Nummer erreichbar ist. Diese radikale Individualisierung der Ansprechbarkeit ist neu, und gilt auch für SMS, E-Mail, Social Networks und vermutlich alles, was noch kommen wird. All diese Instrumente kann und will der "normale" vernetzte Einzelne nebeneinander her benutzen, mit vielen Nuancen in der Verbindlichkeit und Art der jeweiligen Kommunikation. "Reine" Onliner sind demzufolge eine Klischeevorstellung, zu der man in der Wirklichkeit keine Entsprechung findet. (Heller 2008) Für die vernetzte Welt ist vielmehr die komplexe Mischung und das Nebeneinander vieler Medien typisch, die neuerdings jeder einzelne erlernt und benutzt.

“Den” Telefonanschluss oder den Hausbriefkasten, der von einem Familienmitglied bedient wird, gibt es zwar vielleicht noch - aber seine zentrale Bedeutung für das “Netzwerken” hat er verloren.

Die Vernetzung ist auch deshalb so komplex und so individuell geworden, weil sich nicht nur stabile Sets wie Familie oder Nachbarschaft in kleine Netzwerke aufgelöst haben, sondern weil an die Stelle dieser wenigen Sets eine Vielzahl oft spezieller, und typischerweise lockerer, oft nur über Zeitphasen aktiver Verbindungen (“weak ties”) getreten ist. Für diese vielen lockeren Verbindungen spielt räumliche Nähe oft keine Rolle mehr. Das merkt man nicht nur an den Kontakten, die nach dem Auslandssemester im Handy oder in Facebook gespeichert geblieben sind, sondern auch daran, dass Umzüge in eine andere Stadt häufiger werden (übrigens nicht nur in den USA), und sogar daran, dass immer mehr Menschen eine Fernbeziehung führen.

Natürlich spiegelt sich diese Entwicklung, die ich oben mit Beispielen aus Studium und Privatleben veranschaulichen wollte, auch in den typischen Arbeitsbeziehungen wieder. Lose verknüpfte Formen der Zusammenarbeit, die Ersetzung stabiler hierarchischer Beziehungen durch Team- und Freelancer-Tätigkeiten bis hin zur Ausprägung des Arbeitskraftunternehmers (Voß und Pongratz 1998) nehmen zu. Diese Entwicklungen finden in den Netzwerktools auf dem Handy sowie im Internet (was ja zunehmend das gleiche wird) ein passendes Medium.

Last not least erlaubt das Internet auch eine Individualisierung des Informations-Konsums und sogar der -Produktion - im Falle von Facebook werden die Statusmeldungen der eigenen Freunde (oder nüchterner ausgedrückt: der eigenen Netzwerkkontakte) zu demjenigen Medium, in dem am häufigsten gelesen und mit dem am stärksten interagiert wird.

Doch bei Facebook bleibt die Entwicklung nicht stehen - die gleichzeitige Popularisierung weiterer Netzwerke (derzeit Pinterest, Tumblr, spezialisierte Netzwerke zum Beispiel für Wissenschaftler und viele andere) und die auf hohem Niveau stabilisierte Bedeutung der Suchmaschine Google als Navigator in einer Fülle von Informationsquellen und Dokumenten hat zu einer Vervielfachung und Individualisierung des Informationskonsums gesorgt, die historisch beispieldes ist.

Sie ist verbunden mit einem Empowerment der Informationskonsumenten, denen spätestens seit Facebook klar geworden ist, dass nun auch sie selbst öffentliche Informationsproduzenten sein können.

*2. "Here comes everybody" - Ein neues Potential der Organisierung und Artikulation vernetzter Individuen*

Der New Yorker Mediensoziologe Clay Shirky beschreibt, wie auf der von Wellmann und anderen beschriebenen Grundlage einer neuen, individualisiert-vernetzten Medienwelt Interessen artikuliert werden, die in der hoch organisierten Medienwelt vor dem Netzzeitalter "draußen" geblieben waren. In seinem Buch "Here comes everybody - the power of organizing without organizations", erschienen noch vor dem "arabischen Frühling" von 2011, beschreibt Shirky dies anschaulich an der Aufdeckung des jahrzehntelangen sexuellen Kindesmissbrauchs in kirchlichen Einrichtungen. (Shirky 2009)

Man kann, so Shirky, eigentlich nicht sagen, dass das Thema vor den Neunziger nicht bekannt gewesen oder gar in den existierenden Medien bewusst unterdrückt worden wäre. Vielmehr tauchte das Thema an Einzelfällen immer wieder einmal in den Medien auf. Die Kirchen-Entscheider reagierten dann meist, indem sie den jeweiligen Täter als Einzelfall entschuldigten, entließen, oder sogar nur in eine andere Pfarrei versetzten. Die Vielzahl der "Einzelfälle", die zugrunde liegenden Muster, vor allem aber auch der Druck, das Thema allgemein auch in den überregionalen Mainstream-Medien aufzugreifen, das erst zum Einlenken der Kirche, zum Eingeständnis des grundsätzlichen Problems, den Beginn umfassender Entschuldigungen und Entschädigungen führte - dies war, so stellt Shirky schlüssig dar, so nur möglich durch die Vernetzung vieler Einzelner.

Genauer: Die Verbindung zwischen den regionalen Einzelfällen waren durch die Möglichkeiten webbasierter Recherche plötzlich für jedermann einfach nachvollziehbar. Betroffene konnten sich auf einmal sehr viel leichter gegenseitig entdecken. Und sie hatten nun eine Plattform, auf der sie auch ohne eine große Institution, die sich für das Thema "zuständig" betrachtet hätte, kommunizieren, koordinieren, artikulieren und letztlich den entscheidenden öffentlichen Druck entfalten konnten.

Von tiefgreifenden politischen Umwälzungen (die Rolle von SMS, Facebook und Twitter beim "arabischen Frühling" sei hier nur als Stichwort erwähnt) und Medienphänomenen wie die von Shirky beschriebene Artikulation eines gesellschaftlichen Missstandes, auch von banalen Flashmobs (Wikipedia-Autoren 2012) bis hin zur Erfindung des Internets selbst (Johnson 2012) - in den letzten Jahren sind wir immer stärker mit den Auswirkungen des vernetzten Individualismus und seiner Artikulationsformen auf der Grundlage der Netzwerkmedien konfrontiert.

Besonders in einem Bereich werden diese Auswirkungen bereits seit einigen Jahren beklagt und diskutiert: Am schwindenden Bedeutungsverlust traditioneller Medien und Institutionen. So ist in den USA die Zeitungskrise bereits massiv sichtbar, hierzulande dringt sie erst allmählich ins öffentliche Bewusstsein. Und allenthalben klagen Kirchen, Parteien, Verbände und Gewerkschaften über den Verlust an Mitgliedern, vor allem aber auch aktivem Nachwuchs.

*3. In Gestalt heterogener Projekte, die informell von vernetzten Einzelnen (statt von etablierten Institutionen) initiiert werden, tauchen "vernetzter Individualismus" und "Organisierung ohne Organisationen" nun auch in der BID-Branche auf.*

In der Branche der Bibliotheken, Archive, Dokumentations- und Informationseinrichtungen wird seit einigen Jahren diskutiert, wie sich die Vernetzung, siehe oben, auf der Seite der Benutzer – also der Konsumenten der von dieser Branche angebotenen Informations-Produkte und -Dienstleistungen - auswirkt. Wie aus dem einstmals so stabil erschienenen "Benutzer" ein zunehmend flüchtiges Wesen geworden ist - "das Verschwinden des 'Nutzers'", wie Thomas Hapke es in der Überschrift eines Aufsatzes ausgedrückt hat. (Hapke 2007) Bei dieser Dienst- und Output-Orientierung droht der Blick nach Innen zu kurz zu kommen - der Blick auf die sich in Richtung eines vernetzten Individualismus, ubiquitär verfügbarer, sozialer Netzmedien verändernden Strukturen der Aufmerksamkeit, der Teilhabe und des Engagements innerhalb der Informations-Branche selbst.

Werfen wir einen Blick auf einige Ereignisse der letzten paar Jahre.

Im Jahr 2006 wurde von damaligen Studierenden der HU Berlin Libreas gegründet, ein Open Access-Journal für Bibliotheks- und Informationswissenschaften.

Libreas erfreut sich heute großer Beliebtheit bei einigen bekannten Autoren und vielen Lesern in der Bibliothekswelt. (Zum Beispiel Jochum 2009) Die Redakteure arbeiten heute als Informationswissenschaftler und Bibliothekare quer über Deutschland und die Schweiz verstreut.

"Cycling for libraries" ist eine inzwischen jährlich stattfindende Radtour mit z.T. beachtlichem Medienecho in den jeweils bereisten europäischen Ländern, die von einem losen, informellen Netzwerk BibliothekarInnen unterschiedlichen Alters aus unterschiedlichen Bibliothekstypen in vielen verschiedenen Ländern organisiert wird.

Plan3t.info ist ein Portal, in dem mittlerweile mehr als 100 unabhängige Weblogs von BibliothekarInnen, einschließlich berufsverbandlicher BloggerInnen, gebündelt, abonniebar und durchsuchbar sind. Blogs werden typischerweise von Einzelnen geschrieben; unter den in Plan3t aggregierten Blogs befinden sich daneben auch mehrere Blogs, die mit unterschiedlichen Themenschwerpunkten und unterschiedlicher Verbindlichkeit von Gruppen geschrieben werden, z.B. das mittlerweile elf Jahre alte Weblog [log.netbib.de](http://log.netbib.de) oder jüngere Projekte wie [bibliothekarisch.de](http://bibliothekarisch.de) und das Themenblog "Literaturverwaltung und Bibliotheken" ([literaturverwaltung.wordpress.com](http://literaturverwaltung.wordpress.com)).

BRAIN (Potsdam), "Perspektive Bibliothek" (Heidelberg / Potsdam), die Facebook-Gruppe "LIS New Professionals", der Stellen-Suchdienst "Open BiblioJobs", die BibCamps, (Heller 2011) die "Zukunftswerkstatt", frei<TAG> (Berlin / Potsdam), InfoCamp Chur - die Liste ließe sich fortsetzen. All diese Projekte haben in den letzten Jahren eine beachtliche Aufmerksamkeit und vor allem Aktivität aus einem erstaunlich diversen Spektrum der deutschsprachigen bibliothekarischen Öffentlichkeit auf sich gezogen. Die Vielfalt der gewählten Formen ist groß: Es handelt sich um einmalige oder wiederkehrende Veranstaltungen, Diskussionsgruppen, Publikationen, Online-Portale und vieles dazwischen.

Handelt es sich bei den Beteiligten all dieser Projekte um eine relativ homogene, IT-affine Gruppe? Keineswegs. Zur Gründung der oben genannten Facebookgruppe gab es saftige Polemik u.a. von bloggenden BibliothekarInnen zu lesen, die in der Wahl von Facebook als Vernetzungsmedium eine Preisgabe autonomer, zensurresistenter Kommunikation im Netz sahen.

Bei "Cycling for Libraries" wiederum sind zahlreiche Aktive, wie mir Till Kinstler, einer der deutschen Organisatoren, 2012 in einem persönlichen Gespräch berichtete, weder beruflich noch privat sonderlich IT-interessiert. Die BibCamps wiederum haben viele inspirierte Teilnehmer und Nachahmer gefunden, die sich für z.B. für Kernthemen der "Zukunftswerkstatt" wie das Gaming an Öffentlichen Bibliotheken nicht sonderlich interessieren. Die Liste ließe sich fortsetzen, aber zusammengefasst lässt sich festhalten: Diese Projekte sind inzwischen ungefähr genau so vielfältig wie die Landschaft bibliothekarischer Interessen und Aktivitäten selbst.

Und doch hat diese relativ junge, heterogene Projekteszene einen gemeinsamen Nenner: Sie setzt "kooperative Technologien" ein, um es mit einem treffenden Begriff von Christiane Schulzki-Haddouti zu benennen. (Schulzki-Haddouti und Lorenz-Meyer 2008) Diese Projekte sind im Web präsent; sie nutzen es gezielt, um sich und ihre Informationen, Dienste, Produkte, Veranstaltungen bekannter zu machen, Interesse und aktive Teilnahme zu generieren - ohne dabei jedoch für eine etablierte Institution zu sprechen, sei dies nun eine Bibliothek, ein Verbund oder ein Berufsverband.

Letzteres ist der springende Punkt. Die Dynamik dieser Szene fließt zwar als Kraft in die institutionalisierte Bibliothekswelt zurück. So erkennen mittlerweile viele Bibliotheken die Teilnahme an BibCamps als Fortbildung an. Diese Dynamik hat ihren Ausgangspunkt jedoch nicht in jenen Institutionen, sondern im "vernetzten Individualismus" der Beteiligten - diese Dynamik wird durch die Verfügbarkeit neuer Technologien "nur" offen gelegt, verstärkt, in die Arena einer sich wandelnden Öffentlichkeit, in diesem Fall einer bibliothekarischen Fachöffentlichkeit, gebracht - siehe Wellman, siehe Shirky.

*4. In der deutschen BID-Branche herrscht Nachholbedarf in Sachen Vernetzung - auch, weil den neuen Medien in Ausbildung und Beruf bisher oftmals die Anerkennung versagt bleibt.*

Auf den ersten Blick scheint es im Widerspruch zum oben gesagten zu stehen, aber: Die Entfaltung des vernetzten Individualismus in fachlich relevanten Initiativen steckt in der deutschen bibliothekarischen Öffentlichkeit noch in den Kinderschuhen.

Dieser Rückstand lässt sich beispielhaft an aktuellen Zahlen ablesen. OCLC hat 2012 per Email Bibliotheks-MitarbeiterInnen u.a. danach gefragt, auf welchen Wegen sie sich über bibliotheksbezogene Themen informieren. (OCLC) Ich interpretiere insbesondere die Adaption von (Micro-)Blogs, als Medien der selbständigen Vernetzung und Interaktion, als einen starken Indikator für ein Empowerment von Bibliotheks-MitarbeiterInnen im oben beschriebenen Sinne.<sup>1</sup> Aus den 143 bei OCLC eingegangenen Antworten geht hervor, dass erst 28% Blogs und 7% Twitter beruflich verfolgen. Zum Vergleich: In den Niederlanden (152 eingegangene Antworten) liegen die Werte bereits bei 37% (Blogs) und 31% (Twitter).

Ich vermute, dass für dieses Ergebnis unter anderem die verschleppte Modernisierung der Ausbildung der Fachleute in der deutschen BID-Branche verantwortlich ist. Die "kooperativen Technologien" werden, so mein Eindruck, vielfach noch als Orchideen im Glashaus einer kaum verstandenen, fernen Wirklichkeit des Netzes vorgeführt. Sie dürfen zwar hier und da in Abschlussarbeiten auch in ihrer Anwendung "für" oder "in" Bibliotheken untersucht werden. Eine projektartige oder sonstwie praxisnahe Lehre ihrer aktiven Nutzung scheint in Curriculen oder Ausbildungsplänen hingegen kaum verankert zu sei.

Die wenigen Ausnahmen scheinen in der deutschen BID-Landschaft nun unter dem Oberbegriff Innovationsmanagement Anerkennung zu finden.<sup>2</sup>

Es wäre übrigens gut zu wissen, wie viele HochschuldozentInnen, DirektorInnen und PersonalentscheiderInnen die oben genannten Fragen von OCLC nach dem persönlichen Konsum vernetzter Medien für sich selbst bejahen würden. Dem Anschein nach werden es kaum mehr sein als im Gesamtdurchschnitt der Branche. Eine Vorbild- oder Vorreiterfunktion ist hier kaum zu beobachten - eher wird hier der Eindruck verbreitet, dass diese Medien nicht "zählen".

---

<sup>1</sup> Vgl. zur Erschließung des Webs als fachlichen Informationsraum durch bloggende "Amateure" in der Informationsbranche auch die Konzeptstudie, die Heinz Pampel und ich 2010 durchgeführt haben. (Heller und Pampel 2010)

<sup>2</sup> Erwähnenswert sind insbesondere die fachöffentlichen Interventionen Ursula Georgys (FH Köln) zu diesem Thema. (Georgy 2011)

*5. Neugier, Lernbereitschaft und Offenheit für Eigeninitiative innerhalb der Institution ("Intrapreneurship") sind notwendig, um das vernetzte Engagement und Know How in die Institutionen einzubringen und dort zu kultivieren.*

Wenn Sie meinen Prämissen und meiner Bestandsaufnahme bis hierhin gefolgt sind, landen Sie (hoffentlich) auch bei der Frage, wie die bibliothekarischen "Institutionen" mit diesen verschleppten, aber dennoch bereits erkennbaren Veränderungen im Gefüge von Aufmerksamkeit und Engagement der Informationsbranche umgehen.

Eine Irritation, der ich immer wieder in Gesprächen insbesondere mit Mitgliedern der Berufsverbände begegne, ist die Klage über ein zu geringes Interesse der Mitglieder insgesamt, das sich z.B. in geringen Teilnehmerzahlen bei Mitgliederversammlungen ausdrücke. (Das ist nur anekdotisch - ich habe nicht geprüft, ob es einen messbaren Rückgang der Teilnehmerzahlen bei solchen Versammlungen, oder der Kandidaten für Wahlgremien gegeben hat. Die Tatsache, dass so etwas als Problem berichtet wird, kann im Sinne einer beobachtenden Feldforschung jedoch schon an sich aussagekräftig, zumindest interpretierbar sein.)

Was, so könnte man nun im Anschluss an das so beklagte Defizit fragen, ist von der Zerstreuung der Aufmerksamkeit und des Interesses der "Jungen" an zeitweilige, improvisierte, kaum in eine strategisch Verbandspolitik integrierte Mini-Projekte zu halten? Die Antworten darauf sind derzeit noch sehr unterschiedlich. Nach meinem Eindruck durchläuft die Rezeption der oben beschriebenen Entwicklung unterschiedliche Phasen - je nach "Institutionen" und Protagonisten sehe ich Versuche, das Phänomen zu ignorieren (als verschwindende Mode, "rein virtuell" etc.), aber auch Neugier und ganz unterschiedliche Annäherungs- und Adaptionsversuche. (Letzteres leider auch nach dem Motto: Wenn alle auf BibCamps fliegen, dann bezeichne ich als bisher schon erfolgreiche XY-Bibliothek/-Berufsverband meine nächste Veranstaltung einfach als XYcamp...)

Die Neugier, so ist zu hoffen, bleibt. Dauerhaft und erfolgreich kann sie m.E. nur werden, wenn die für und in den Institutionen Handelnden den vernetzten Individualismus zunächst als für sich selbst relevante Ressource entdecken, sich mit den Medien vertraut machen, und sich auf den Weg machen, sich den bestehenden Communities mit Offenheit und Lernbereitschaft zu nähern.



Denn die Formate und Techniken selbst sind bekanntlich längst nicht mehr die entscheidende Herausforderung. Heute kann sich jeder binnen weniger Minuten Blogs und Wikis einrichten, oder seine Veranstaltungen als Barcamp oder BibCamp bezeichnen. Umso schwieriger und interessanter ist jedoch die Frage geworden, unter welchen Voraussetzungen ein solches Blog-, Wiki-, Barcamp- etc. Projekt die rare Aufmerksamkeit potenzieller Rezipienten, geschweige denn das Vertrauen potenzieller "Mitmacher" gewinnt.

Dass diese Neugier bleibt und zum aktiven Lernen wird ist wichtig, weil "Netzwerk" und "Institution" voneinander profitieren können. Wer bei der DGI-Konferenz 2012 den Vortrag von Adam Sofronijevic (Universität Belgrad) hörte, mag überrascht gewesen sein, wie manche Institutionen heute gezielt die vernetzte Selbständigkeit ihrer MitarbeiterInnen freisetzen, auch als Bibliothek. (Sofronijevic 2012) Bei Sofronijevic fällt das unter das Stichwort Intrapreneurship, ein Kompositum aus intracorporate (unternehmensintern) und Entrepreneurship (Unternehmertum).

So haben etwa vom - personell, finanziell und zeitlich begrenzten - Engagement der Unterstützung eines BibCamps bereits einige deutsche Bibliotheken und Hochschulen profitiert: Als Arbeitgeberin oder Ausbilderin im Wettbewerb um qualifizierte und engagierte Köpfe ("Employer Branding"), aber auch als Institution, die sich zukunftsrelevantes Know How ins Haus holt, und sei es zunächst nur zwei turbulente Tage lang.

Sofronijevic verrät übrigens auch einen Trick, Intrapreneurship dieser Art zu entzünden. BibCamps und andere Aktivitäten dieser Art lassen sich ihm zufolge nur schwer "top down" anordnen - es funktioniert besser, vorgefundene "Grassroot"-Aktivitäten zu unterstützen. Manchmal reicht es schon, dem vernetzten Individualismus dort mit Offenheit und Lernbereitschaft zu begegnen, wo er sich im Umfeld der eigenen Institution bereits zeigt.

## Literaturverzeichnis

- Georgy, Ursula. 2011. Open Innovation - Integration der Hochschulen in den Innovationsprozess von Bibliotheken. In: 100. Deutscher Bibliothekartag. Berlin: BIB. <http://www.opus-bayern.de/bib-info/volltexte/2011/1083/> (Zugegriffen: 29. September 2012).
- Hapke, Thomas. 2007. Informationskompetenz 2.0 und das Verschwinden des „Nutzers“. *BIBLIOTHEK Forschung und Praxis* 31, Nr. 2: 137–149. <http://hdl.handle.net/10760/10530> (Zugegriffen: 29. September 2012).
- Heller, Lambert. 2008. Wie sozial ist das Internet? *Biblionik*, 30. August. <http://biblionik.de/2008/08/30/wie-sozial-ist-das-internet/> (Zugegriffen: 29. September 2012).
- . 2011. Blogs, BibCamps, Lightning Talks – bibliothekarische Kommunikation in Bewegung. *Goethe-Institut e.V.*, Abschn. Bibliotheken in Deutschland – Fachdiskussion. <http://www.goethe.de/wis/bib/fdk/de8569828.htm> (Zugegriffen: 29. September 2012).
- Heller, Lambert und Heinz Pampel. 2010. Konzeptstudie: Die informationswissenschaftliche Zeitschrift der Zukunft. In: *Semantic Web & Linked Data : Elemente zukünftiger Informationsinfrastrukturen*; 1. DGI-Konferenz, 62. DGI Jahrestagung. Frankfurt am Main: DGI. <http://eprints.rclis.org/19033/>.
- Jochum, Uwe. 2009. Der Souverän. *Libreas - Library Ideas* 15 (Frühjahr). <http://edoc.hu-berlin.de/docviews/abstract.php?id=30240> (Zugegriffen: 29. September 2012).
- Johnson, Steven. 2012. The Internet? We Built That. *The New York Times*, 21. September, Abschn. Magazine. <http://www.nytimes.com/2012/09/23/magazine/the-internet-we-built-that.html> (Zugegriffen: 29. September 2012).
- OCLC. Libraries in Germany: Priorities & Perspectives. <http://www.oclc.org/de/de/reports/de-libraries/> (Zugegriffen: 29. September 2012).
- Schulzki-Haddouti, C. und L. Lorenz-Meyer. 2008. *Kooperative Technologien in Arbeit, Ausbildung und Zivilgesellschaft: Analyse für die Innovations-und Technikanalyse (ITA) im Bundesministerium für Bildung und Forschung (BMBF) im Rahmen eines Forschungsprojekts am Fachbereich Media der Hochschule Darmstadt 2008*. Hochschule Darmstadt.
- Shirky, Clay. 2009. *Here Comes Everybody: The Power of Organizing Without Organizations*. Reprint. Penguin Books.
- Sofronijevic, Adam. 2012. Enterprise 2.0 in a library - European perspective. In: *Social Media and Web Science : das Web als Lebensraum*; 2. DGI-Konferenz, 64. Jahrestagung der DGI. Düsseldorf: DGI.
- Voß, G. und H. G. Pongratz. 1998. Der Arbeitskraftunternehmer. Eine neue Grundform der Ware Arbeitskraft? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 50: 131–158.
- Wellman, B., A. Quan-Haase, J. Boase, W. Chen, K. Hampton, I. Díaz und K. Miyata. 2003. The social affordances of the Internet for networked individualism. *Journal of Computer-Mediated Communication* 8, Nr. 3. <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2003.tb00216.x/full> (Zugegriffen: 29. September 2012).
- Wikipedia-Autoren. 2012. Flashmob. *Wikipedia*. 25. August. <http://de.wikipedia.org/w/index.php?title=Flashmob&oldid=107244817> (Zugegriffen: 29. September 2012).

---

## **Kollaborative Wissensgenerierung im virtuellen Raum – Entwicklungen und Erkenntnisse aus der internationalen Bildungsmedienforschung**

Sylvia Brink, Andreas L. Fuchs, Roderich Henry, Robert Strötgen

Georg-Eckert-Institut für internationale Schulbuchforschung, Braunschweig

### **Zusammenfassung**

Virtuelle Forschungsumgebungen haben in den letzten Jahren in nahezu allen Bereichen der Wissenschaft Einzug gehalten. Auch in der internationalen Bildungsmedienforschung entsteht mit Edumeres.net ein Portal, das kollaborative Wissensgenerierung im virtuellen Raum ermöglichen soll. Der vorliegende Beitrag dokumentiert und erläutert die Erfahrungen aus der Entwicklung und den ersten Jahren des Betriebs sowie die daraus resultierenden Konsequenzen für eine zukünftige Vernetzung von Daten und Menschen.

### **Abstract**

In the last few years virtual research environments have become part of nearly every field of scientific research. One of these platforms is Edumeres.net, developed for virtual collaboration in international educational media research. This article illustrates and comments on the experiences made in the first years of operation and the resulting consequences for further developments and for future networking of data and people.

Am 1. April 2004 startete eines der ersten Förderprogramme zur Etablierung von Virtuellen Forschungsumgebungen (VFU).<sup>1</sup> Die fördernde Organisation, das *Joint Information Systems Committee*, kurz *JISC*, beheimatet in Großbritannien, verwendete dabei den Begriff der Virtual Research Environment, der sich im Laufe der Zeit, so weit durchsetzte, dass auch in deutschen Publikationen immer wieder das Akronym VRE<sup>2</sup> zu finden ist. Was sich dahinter verbirgt ist jedoch bis heute nicht eindeutig definiert.

---

<sup>1</sup> Die erste Phase des „Virtual Research Environments programme“ des JISC dauerte vom 1. April 2004 bis zum 31. März 2007. Daran schlossen sich noch zwei weitere Förderphasen an. Genauere Angaben dazu finden sich unter <http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>.

<sup>2</sup> „Apart from the term 'VRE', these are some of the related English terms used: Collaborative e-Research Communities, Collaborative Virtual Environment, Collaboratory, (Science) Gateway, Virtual Organisation, Virtual Research Community.“ (Carusi, Reimer 2010, 13)

Der vorliegende Beitrag greift diese Entwicklung auf und verknüpft sie mit der Konzeption und Umsetzung des Informations- und Kommunikationsportals Edumeres.net. Zunächst wird näher auf den Begriff VRE eingegangen. Anhand der Analyse von Definitionen aus internationalen Programmen zum Aufbau von Forschungsumgebungen zeigt sich, dass dieser einen breiten Interpretationsspielraum zulässt. Um in solch einem Fall der Beliebigkeit zu entgehen, wird, in Anlehnung an die Intermedialität der Begriff des Kommunikationsdispositivs eingeführt und zwischen einer engen und weiten Sichtweise von VRE unterschieden. Vor dem Hintergrund der sich daraus ergebenden Konsequenzen wird die Planung, die technische Einrichtung und der bisherige Betrieb von Edumeres.net kritisch beleuchtet und mit den User-Evaluationen und den in den ersten zwei Jahren gewonnenen Erfahrungen in Zusammenhang gesetzt. Abschließend steht der Ausblick auf das weitere Vorgehen bei der Vernetzung der internationalen Bildungsmedienforschung durch Edumeres.net.

### **Virtuelle Forschungsumgebung (VRE) – Ein weites Feld**

Am Anfang jeder Idee steht eine Vision. So auch beim einleitend genannten *JISC* Programm zur Etablierung von Virtuellen Forschungsumgebungen. Auf dessen Internetseite<sup>3</sup> folgt gleich nach der Nennung der drei Förderphasen ein Absatz, überschrieben mit „Vision“. Darunter findet sich neben Anderem eine knappe Definition von „VRE“:

“A VRE helps researchers from all disciplines to work collaboratively by managing the increasingly complex range of tasks involved in carrying out research.” (JISC 2012)

In nahezu gleichem Wortlaut wurde diese Formulierung auch in die *JISC-Landscape-Studie* von 2010 übernommen und erweitert:

“The term VRE is now best thought of as shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers (who may come from different disciplines, institutions or even countries) and to make use of resources and technical infrastructures available both locally and nationally.” (Carusi, Reimer 2010, 13)

---

<sup>3</sup> <http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>

Gekürzt auf einige Schlagworte ergibt sich folgender Sachverhalt: Eine Virtuelle Forschungsumgebung beinhaltet Werkzeuge, Ressourcen und Technologien, die es Wissenschaftlerinnen und Wissenschaftlern ermöglichen, kollaborativ über Disziplin- und Ländergrenzen hinweg zu forschen.

Welche Technologien genauer dahinter stehen wird in dieser Definition allerdings nicht näher ausgeführt. Somit könnten beispielsweise auch Telefon- und Faxgeräte einer VRE zugerechnet werden. Diese offene Formulierung ist aber durchaus gewollt:

“The definition is fairly vague and deliberately so. [...] The definition reflects the fact that the concept of a VRE is indeed evolving and all representatives of funding bodies we spoke to agree that being too prescriptive in the definition might stifle innovation.” (Carusi, Reimer 2010, 13)

Auch andere Institutionen und Untersuchungen scheinen einer freien Entwicklung keine Steine durch restriktive Festlegungen in den Weg legen zu wollen.

Die ebenfalls im Jahr 2010 erschienene Studie: *Collaboratories: Connecting Researchers* der SURFfoundation in Utrecht hält Ihre Definition ebenfalls sehr knapp:

“A Collaboratory, or virtual research workplace/environment, is a web-based collaboration environment for researchers. The literature also describes it as: ‘an organisational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artifacts and tools required to accomplish research tasks.’” (Vaart 2010, 13)

Was in diesem Fall besonders ins Auge fällt ist das Wort “web-based”. Im Gegensatz zu anderen Stellen in der Literatur, die möglicherweise wegen der Bezeichnung „virtuell“ von vornherein davon ausgehen, wird hier eine Bindung an das Internet explizit mit in die Begriffsbestimmung aufgenommen. Die *Allianz der deutschen Wissenschaftsorganisationen* wiederum verzichtet in ihrer 2011 veröffentlichten VRE-Definition darauf:

„Eine virtuelle Forschungsumgebung ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse - während sie

technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert.“ (Allianzinitiative 2011)

Auch hier liegt die Betonung auf kollaborativem Arbeiten zwischen mehreren Wissenschaftlern über größere Distanzen hinweg. Gegenüber den anderen Definition kommt dabei noch eine Erklärung hinzu, was unter einem Forschungsprozess zu verstehen sei: die Erhebung, die Diskussion und die weitere Bearbeitung von Daten sowie die Publikation von Ergebnissen. In den bisherigen Beschreibungen wurde vornehmlich von „research tasks“ oder ganz allgemein von „research“ gesprochen. Dennoch erscheint auch bei diesen ergänzenden Angaben der Begriff VRE von seiner Funktion her sehr weit gefasst.

Des weiteren bleibt offen wer oder was eine Virtuelle Forschungsumgebung ausmacht. Ist es ein rein technisches Verständnis, das sich auf die IT-Infrastruktur bezieht oder schließt es das ganze Forschungsszenarium mit ein, dem auch die einzelnen Forscher und Institutionen zuzurechnen sind?

Die Arbeitsgruppe "Virtuelle Forschungsumgebungen (vForum)" der *Deutschen Initiative für Netzwerkinformation (DINI)* sieht es folgendermaßen:

„Der Begriff ‚Virtuelle Forschungsumgebung‘ beschreibt dabei ein breites Spektrum an IT-gestützten Forschungsszenarien, das von der Kommunikation in der Gruppe über das verteilte Schreiben bis zur gemeinsamen Durchführung von Experimenten und den Zugriff auf tausende Kilometer entfernte Geräte und Sensoren reicht.“ (DINI 2012)

Im Zentrum steht also hier das Forschungsszenarium. Die Technik, das heißt die IT-Infrastruktur ist unterstützender, aber nicht alleiniger Teil einer VRE. Der Vorgang des Kommunizierens, scheint hier ebenso dazuzugehören wie das Medium mit dem jene zustande kommt. Dies wird bekräftigt durch folgenden, sich an obige Textstelle nahtlos anschließenden Satz:

„Über die wissenschaftlichen Inhalte hinaus spielen vor allem die Aspekte ‚wie‘ [...] und ‚was‘ [...] eine zentrale Rolle bei der Entwicklung von IuK-Dienstleistungen für Virtuelle Forschungsumgebungen.“ (DINI 2012)

Es wird zwar über die Entwicklung von Informations- und Kommunikations-Dienstleistungen gesprochen, doch deutet das Wort „für“ an, dass diese nicht alleine die VRE darstellen.

Eine etwas andere Perspektive zeigt die oben schon aufgeführte *Allianz der deutschen Wissenschaftsorganisationen*. Sie schreibt in ihrer Definition:

„Virtuelle Forschungsumgebungen sind wesentliche Komponenten moderner Forschungsinfrastrukturen.“ (Allianzinitiative 2011)

Hier ist die Blickrichtung gewendet. Virtuelle Forschungsumgebungen sind Teil der Forschungsinfrastruktur. Die Aussage: „[e]ine virtuelle Forschungsinfrastruktur ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit [...] ermöglicht“ (Allianzinitiative 2011), deutet darauf hin, dass das Verständnis einer VRE hier eher in eine technische Richtung geht. Inhaltlich sprechen die *DINI* und die *Allianz* von den selben Dingen, einzig die Begriffe werden unterschiedlich verwendet.

Ähnliche Gemeinsamkeiten finden sich auch zwischen dem an das *JISC*-Förderprogramm angelehnten Artikel von Fraser von 2005 und dem bereits 2003 erschienenen Atkins-Report:

“Between these two layers is the cyberinfrastructure layer of enabling hardware, algorithms, software, communications, institutions, and personnel.” (Atkins et al. 2003)

Hinter dem vor allem in den USA verwendeten Begriff der “cyberinfrastructure” verbirgt sich bei Atkins ähnlich der *DINI*-Definition ein weitaus breiteres Verständnis einer VRE, die namentlich auch Institutionen und Menschen mit einschließt. Diesem stimmt auch Michael Fraser zu:

“For the most part, the terms VRE, cyber- or e-infrastructure are synonymous. If there is a difference between a VRE and cyber/e-infrastructure then it is this: the VRE presents a holistic view of the context in which research takes place whereas e-infrastructure focusses on the core, shared services over which the VRE is expected to operate. A VRE is more than middleware [sic! ...].” (Fraser 2005)

Virtuelle Forschungsumgebungen sind auch bei ihm größere Einheiten, die über Hard- und Software hinausgehen. Würde der erste Satz für sich stehen, bestände auch kein Unterschied zu Atkins. Erst die Differenzierung der Begriffe VRE und cyberinfrastructure macht eine unterschiedliche Verwendung der Terminologie sichtbar.

Wie ist nun mit einer solchen Situation umzugehen, in der weltweit über die selben Dinge gesprochen wird, deren Inhalt jedoch zum einen nur vage und offen definiert ist, zum anderen keine eindeutige Terminologie vorliegt? Eine genauere Festlegung



hätte sicherlich zur Folge, dass beispielsweise Forschungsverbünde leichter in Austausch treten könnten, wenn alle beteiligten Partner die selben Vorstellungen hätten. Des weitern könnten Förderprogramme gezielter ausgeschrieben werden. Eine Fort- und Weiterentwicklung von Virtuellen Forschungsumgebungen würde dadurch allerdings in seiner Vielfältigkeit einschränkt.

Einen möglichen Ausweg könnten Anleihen aus der Medienwissenschaft bzw. Intermedialitätsforschung bieten.<sup>4</sup> Der Begriff des Mediums wird dort in einer noch viel breiteren Bedeutungsvielfalt gebraucht als der Terminus VRE in den Förder- und Forschungsprogrammen der Wissenschaft, dennoch lassen sich Parallelen ziehen.

Werner Wolf schreibt in seinem Buch *Musicalization of Fiction*:

“As critics have pointed out, there is a plethora of meanings connected with the word 'medium', meanings that only partially converge and range from a narrow to an at times extremely broad definition.” (Wolf 1999, 35)

Das gleiche trifft, wie oben beschrieben, auch auf die Interpretation Virtueller Forschungsumgebungen zu: Sie reicht von einer engen, rein technischen Betrachtung hin zu einer weiten, das gesamte Forschungsszenario einschließenden Sichtweise.

Im Falle des Begriffs Medium findet Wolf deshalb in seinem späteren Aufsatz „Intermedialität: Ein weites Feld und eine Herausforderung für die Literaturwissenschaft“ eine eigene Definition:

„[...] bedeutet 'Medium' hier [...] nicht vorrangig einen bloß technisch-materiell definierten Übertragungskanal von Informationen (wie z.B. Schrift, Druck, Rundfunk, CD usw.), sondern ein konventionell im Sinn eines kognitiven *frame of reference* als distinkt angesehenes Kommunikationsdispositiv.“ (Wolf 2002, 165)

---

<sup>4</sup> Der hier folgende Vergleich mag aus Sicht der Intermedialitätsforschung und Literaturwissenschaft sehr kurz erscheinen und viele Fragen offen lassen, dennoch sei versucht innerhalb des knappen Rahmens dieses Beitrags einige Gedanken anzustellen, die an anderer Stelle auszubauen sind. So würde es sich auch anbieten, Virtuelle Forschungsumgebungen näher aus dem Gesichtspunkt eines Mediums zu betrachten oder in diesem Zusammenhang diskurstheoretische Überlegungen anzustellen. Nimmt man ein Diskurskonzept, welches "Kommunikation sowie die Entstehung, Zirkulation und Distribution von Wissen als kontingente Effekte 'überindividueller', sozial strukturierter Praktiken" (Keller et al. 2005, 8) ansieht, so öffnen sich hier Wege, die ebenso weiterzudenken sind.

Auch auf dem Gebiet der Informationswissenschaft wird in diese Richtung argumentiert: "In den vergangenen Jahren wurde zunehmend versucht, den humanorientierten und den technologieorientierten Ansatz miteinander zu verknüpfen. Dafür sollen die kreativen und intellektuellen Fähigkeiten eines Individuums beim Umgang mit Wissen, und die daten- und informationsverarbeitenden Kapazitäten der Computertechnologie miteinander verbunden werden, um Synergieeffekte zu erzielen." (Gneiting 2011, 12-13)

Übertragen auf die Frage nach dem Verständnis von VRE würde dies heißen, eine Virtuelle Forschungsumgebung stellt ein Kommunikationsdispositiv dar.

Der erste Wortteil, die Kommunikation steht zweifellos außer Frage. Eine der wichtigsten Voraussetzungen für Wissenschaft und damit für alle Formen von Forschungsumgebungen unerlässlich ist die Kommunikation. Ohne Austausch entsteht nur spärlich neues Wissen.

Für den Umfang dieser Kommunikation steht nun der zweite Wortteil, der von Michel Foucault entwickelte Begriff des Dispositivs.

„Das was ich mit diesem Begriff zu bestimmen versuche, ist [...] eine entschieden heterogene Gesamtheit, bestehend aus Diskursen, Institutionen, architektonischen Einrichtungen, reglementierenden Entscheidungen, Gesetzen, administrativen Maßnahmen, wissenschaftlichen Aussagen, philosophischen, moralischen und philanthropischen Lehrsätzen, kurz: Gesagtes ebenso wie Ungesagtes, das sind die Elemente des Dispositivs. Das Dispositiv selbst ist das Netz, das man zwischen diesen Elementen herstellen kann.“ (Foucault 2008, 239)

Zusammengesetzt bedeutet dies für den Begriff der Virtuellen Forschungsumgebung, dass sie das Netz darstellt, das alle Beteiligten der wissenschaftlichen Wertschöpfung miteinander verbindet. Dazu zählen Forscher und Forschungseinrichtungen, eine interessierte Öffentlichkeit, Drittmittelgeber und Nachnutzer gewonnener Ergebnisse, die Hard- und Software, die die Kommunikation erst ermöglicht und alle Prozesse, die zwischen diesen stattfinden. Anders ausgedrückt, „[n]ach der gegebenen Definition ist ein Dispositiv [- in diesem Fall eine Virtuelle Forschungsumgebung -] eine Gesamtheit von Institutionen, Diskursen und Praktiken.“ (Ruoff 2007, 101)

Zugegeben, die Verknüpfung von VRE und Kommunikationsdispositiv führt zu keiner enger gefassten und damit scheinbar präziseren Definition der Inhalte einer Virtuellen Forschungsumgebung, sondern erweitert diese. Doch gerade dadurch kann sie den Austausch darüber erleichtern, da über die Multiperspektivität ein Begriff eingeführt wird, der sich nicht auf spezielle Vorstellungen beschränkt, sondern vieles einschließt, was sonst möglicherweise unbedacht bliebe. Für konkrete Situationen bietet sich dann, angelehnt an Werner Wolf, eine Unterscheidung nach einem engen und einem weiten VRE-Verständnis an.

Erstgenannter setzt bei einer rein technischen Ausrichtung an, die sich auf kollaborative Software im Sinne des Web 2.0 konzentriert, letztgenannter geht über einen Einbezug von Datenbanken, virtuellen Forschungsbibliotheken und Publikationsrepositorien hin zu einem weiten Verständnis, das den gesamten Prozess der Wissensgenerierung inklusive der forschenden Akteure einschließt. Dies steckt einen Rahmen ab, begrenzt aber nicht. Denn „das Dispositiv ist nicht einfach gegeben, sondern es entwickelt sich erst unter der Vorgabe seiner Funktion in der historischen Anordnung.“ (Ruoff 2007, 101)

Wie dies im konkreten Fall aussehen kann, soll nun anhand des Informations- und Kommunikationsportals Edumeres.net gezeigt werden, dessen Konzeption sich im Laufe der letzten Jahre von einem engen zu einem weiten Verständnis entwickelte.

### **Die Entwicklung der Virtuellen Forschungsumgebung (von) Edumeres.net**

Die ersten Überlegungen am *Georg-Eckert-Institut* zur Etablierung einer Internetplattform für die Bildungsmedienforschung reichen etwa fünf Jahre zurück. Einer der Motivationsgründe dafür findet sich auf der Homepage des Instituts:

„Die zentrale Aufgabe des Georg-Eckert-Instituts besteht in der Erforschung historisch, politisch und geographisch bedeutsamer Darstellungen in Schulbüchern und anderen schulrelevanten Bildungsmedien. [...] Schulbuchforschung ist ein multidisziplinär strukturiertes Querschnittsfeld, das als akademische Disziplin an keiner Universität zu finden ist. Da dieses Feld aber für Wissenschaftler aus unterschiedlichen disziplinären Kontexten und Regionen attraktiv und anschlussfähig ist, sind sie auf ein „physisches Zentrum“ und ein forschungsförderndes Umfeld angewiesen.“ (GEI 2012)

Über viele Jahre hinweg bildete das Institut mit seiner Bibliothek in Braunschweig dieses Zentrum. Mit den wachsenden Möglichkeiten des Internets sollte dieser Auftrag nun seine Entsprechung im Netz finden, sollte die Bildungsmedienforschung auch virtuell stattfinden. Ausgangspunkte für den Einsatz einer Virtuellen Forschungsumgebung waren durch das Fächer- und geographische Grenzen überschreitende Forschungsfeld gegeben.

### **VRE – Entwicklung und Betrieb**

In einem ersten Schritt wurde mit Hilfe des Content Management Systems Typo3 ein Webaufttritt erarbeitet, der im Februar 2009 zum *DINI*-Workshop „Förderung der wissenschaftlichen Informationslandschaft in Deutschland“<sup>5</sup> online ging. Auf dieser Konferenz wurden Projekte vorgestellt, die in der ersten Runde der *DFG*-Ausschreibung eine Förderungszusage erhalten hatten. Auch Edumeres.net gehörte zu dieser Gruppe und hatte nun zwei Jahre Zeit, das Konzept des Antrags umzusetzen. Dieses sah vor die so eben erstellte Informationsplattform durch eine Virtuelle Forschungsumgebung zu ergänzen. Abbildung 1 (siehe unten) zeigt den Stand vom Frühjahr 2009. Alle Navigationspunkte bieten Informationen an, von aktuellen Beiträgen, über Schulbuchrezensionen und -analysen hin zu einem Glossar. Die Navigationspunkte mit zwei Icons ,weisen darauf hin, dass der Nutzer hier die Möglichkeit hat, aktiv zu werden, sei es über die Kommentarfunktion oder eigene Beiträge. Diese wurden damals jedoch noch am eigenen PC erstellt und per Email an die Redaktion gesandt. Eine Forschungsumgebung sollte nun im Rahmen der *DFG*-Förderung virtuelle Möglichkeiten dafür schaffen.

---

<sup>5</sup> Der gemeinsame Workshop der *Deutschen Forschungsgemeinschaft*, Gruppe „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ und die Arbeitsgruppe „Elektronisches Publizieren“ der *Deutschen Initiative für Netzwerkinformation (DINI)*, fand am 17. und 18. Februar 2009 in Berlin statt und stand unter dem Titel „Förderung der wissenschaftlichen Informationslandschaft in Deutschland - Der Aufbau Virtueller Forschungsumgebungen.“

edumeres.net Informations- und Kommunikationsportal Internationale Bildungsmedienforschung

PDF-Version Druckversion

Startseite  
Bildungsmedienkataloge  
Literaturrecherche  
Rezensionen  
Analysen  
Bildungssysteme  
Glossar  
Publikationen

**Bildungssysteme** Bildungs- und Lehrpläne Schulbuchsysteme

**Bildungssysteme**

Graphische Übersichten zu den Bildungssystemen Frankreichs, Italiens, Deutschlands und der Niederlande (von links nach rechts):

Unter folgenden externen Links finden Sie weitere Informationen zu nationalen und internationalen Bildungssystemen:

Bildungssysteme in **Deutschland** auf dem Deutschen Bildungsserver:  
<http://www.bildungsserver.de/zeigen.html?seite=505>

Informationsnetz zum Bildungswesen in **Europa**:  
<http://eacea.ec.europa.eu/portal/page/portal/Eurydice>

Datenbank zu den Bildungssystemen in **Europa**:  
<http://eacea.ec.europa.eu/portal/page/portal/Eurydice/EuryPresentation>

**World Data of Education** sixth edition (2006/2007) (UNESCO):  
<http://www.ibe.unesco.org/Countries/WDE/2006/index.html>

Abbildung 1: Screenshot von Edumeres.net (Stand: 18.02.2009)

Ganz im Sinne des Web 2.0 wurden von einem sechsköpfigen Redaktionsteam interaktive Komponenten für die Internetplattform erprobt, in Zusammenarbeit mit einem Typo3-Dienstleister angepasst, gegebenenfalls neu erstellt und auf einem Testserver zu einem einheitlichen System, das später über eine Login-Funktion zu erreichen sein sollte, zusammengesetzt. Um herauszufinden welche Funktionen Wissenschaftlerinnen und Wissenschaftler in der Bildungsmedienforschung benötigen, wurden die Mitarbeiterinnen und Mitarbeiter der wissenschaftlichen Arbeitsbereiche des Instituts interviewt. Etliche Extensions für Typo3 durchliefen User-Evaluationen bis sich neben einer Projektdaten- und Mitgliederverwaltung (Nr. 1 in Abbildung 2) ein Set von vier Werkzeugen herauskristallisierte: ein Forum zur projektinternen Kommunikation (2), ein Textwerkzeug zum kollaborativen Verfassen von Texten (3), ein Dateimanager (4) und eine Blogfunktion (5) für die externe Kommunikation. Die größten Schwierigkeiten bereitete dabei das Tool zur Texterzeugung. Es sollte gleichzeitiges Arbeiten an einem Dokument ermöglichen, dabei aber über eine History-Funktion verfügen, die jederzeit nachvollziehen lässt,

welcher Nutzer wann was geändert hat und somit auch ein Wiederherstellen älterer Versionen zulässt. Am wichtigsten jedoch war, dass das Aussehen und die Bedienung so nah wie möglich an die von den meisten Nutzern gewohnte Word-Umgebung aus dem MS-Office Programm heran reicht. Dadurch schied ein Wiki-System zum damaligen Zeitpunkt aus. Nach einem Versuch mit einem eigenen Texteditor auf Basis von Typo3 fiel die Wahl auf den Online-Office-Anbieter Zoho, der am nächsten an die Anforderungen herankam. Sobald einzelne Komponenten funktionsfähig waren, wurden sie in das Live-System übertragen. Ausgewählte Nutzer konnten sich so bereits registrieren und die Redaktion konnte so die Forschungsumgebung im Betrieb testen.

**edumeres.net** Educational Media Research

DE / EN

afuchs@gei.de ist eingeloggt. Logout

Informationen 6 Publikationen 7 Themenzugänge 8 Virtuelle Forschungsumgebung 9

**PROJEKTBEREICH**

- Meine Projekte
- Alle Projekte
- Neues Projekt

**Forschungsinfrastrukturen**

AG virtuelle Forschungsinfrastrukturen

**Zusammenfassung**

AG virtuelle Forschungsinfrastrukturen

**Metadaten**

Sichtbarkeit: Eingeschränktes Projekt

Projektstatus: Idee

angelegt am: 09.03.2011

letzte Änderung: 04.04.2011

Mitgliedstatus: Leiter

Projektleiter: afuchs@gei.de

**Letzte Änderungen**

Seite: 1 2 > >>

**Mitglieder**

Name	E-Mail	Rolle	Letzte Änderungen
Brink, Sylvia	brink@gei.de	Mitglied	09.03.2011, 12:24
Fuchs, Andreas L.	afuchs@gei.de	Projektleiter	09.03.2011, 12:18
Henry, Roderich	henry@gei.de	Mitglied	09.03.2011, 12:23
Reiß, Kathleen	reiss@gei.de	Mitglied	09.03.2011, 12:23

**PERSONLICHER BEREICH**

- Mein Profil
- Profil bearbeiten

**KONTAKTBEREICH**

- Meine Kontakte
- Alle Kontakte
- Kontakt Suche

**11 NEUE NACHRICHTEN**

Nachrichten einsehen

**Offene Kontaktanfragen**

Annahmen... Ablehnen...

Abbildung 2: Screenshot der VRE von Edumeres.net (Stand: 01.08.2012)

### **VRE – Erfahrungen aus zwei Jahren Betrieb**

Am 19. Februar 2011 erfolgte die offizielle Freischaltung der Virtuellen Forschungsumgebung. Bis heute haben sich 310 Nutzer registriert, die insgesamt 81 Projekte anlegten. Die Bandbreite reichte dabei von einzelnen Wissenschaftlerinnen und Wissenschaftlern, die die VRE vor allem als Online-Schreib-Tool und Dateiablage nutzten bis hin zu Seminargruppen, die kollaborativ Texte erstellten, über das Forum kommunizierten und Seminarunterlagen austauschten. Im Moment sind 35 Projekte offen, von denen etwa 10 in letzter Zeit aktiv betrieben wurden.

In diesen etwa zwei Jahren des produktiven Betriebs stellten sich Erfahrungen ein, die nur über einen längeren Zeitraum gewonnen werden konnten. Die Usabilitytests waren sinnvoll, um die Technik und die Anwendung zu prüfen; die Annahme oder Ablehnung des Angebots der VRE von Edumeres.net durch den Nutzer zeigte sich erst im laufenden Betrieb. Erwartungsgemäß kam das meiste Feedback zum Zoho Writer, weckte er doch mit der Möglichkeit gleichzeitig an einem Text schreiben zu können die meiste Aufmerksamkeit. Wie sich allerdings zeigte, befassten sich viele Nutzer, die gemeinsamen an einem Text arbeiteten, zu unterschiedlichen Zeiten mit dem Dokument. Wichtiger war ihnen daher die Möglichkeit der Änderungsnachverfolgung, was in der gewählten Version von Zoho Writer, bei der die Dokumente auf einem eigenen Edumeres.net-Server verbleiben, zur Zeit nicht möglich ist. Am häufigsten in den Projekten genutzt wurde der Dateimanager, der zusammen mit dem Forum ohne größere Probleme bei den Projektmitgliedern zur Anwendung kam. Am seltensten zum Einsatz kam die Blog-Funktion zur externen Kommunikation, was vermutlich daran lag, dass nahezu alle Projekte einen geschlossenen User-Kreis hatten.

Neben der Verwendung der einzelnen Werkzeuge galt besonderes Interesse der Frage, welche Projekte die größte Aktivität aufwiesen. Während solche, die nur von einem oder zwei Nutzern betrieben wurden, eine kurze Lebensdauer hatten und meist ohne Ergebnis blieben, wiesen Gruppenprojekte größere Aktivität auf. Am erfolgreichsten waren Projekte, die als Ziel eine Publikation auf Edumeres.net hatten. Diese standen meistens auch im intensivsten Kontakt mit der Redaktion.

Als Beispiel lassen sich hier die Rezensionenprojekte nennen, bei denen Studentengruppen in einem Seminar Schulbücher analysiert und gemeinsam Rezensionstexte in der VRE verfasst haben.

Am Ende der jeweiligen Semester wurden daraus eine mit URN versehene und bei der Deutschen Nationalbibliothek gemeldete Veröffentlichungen, die im Publikationsteil von Edumeres.net eingestellt wurden.

#### **VRE – Konsequenz und Zukunft: Edumeres.net als Kommunikationsdispositiv**

Welche Konsequenzen wurden und werden nun aus diesen Erfahrungen für die weitere Entwicklung von Edumeres.net gezogen? Bereits 2011 führten die ersten Erkenntnisse dazu, dass sich der Blick von einem anfänglich engem, rein auf die technischen Möglichkeiten einer Virtuellen Forschungsumgebung konzentrierten Sichtweise geweitet hat zu einem das gesamte Portal umfassende Verständnis. Das heutige Design von Edumeres.net (Abbildung 2) gibt noch Aufschluss darüber, dass die VRE (9) zunächst nur als eine von vier Teilen gesehen wurde. Zwar gab es bis jetzt am Layout soweit keine Änderungen, doch gründen seitdem die weiteren Arbeiten an der Plattform auf einem alle Inhalte und Funktionen von Edumeres.net einschließenden Konzept. Durch die Einbindung von Recherchemöglichkeiten (6 + 8) wie Datenbanken und virtuellen Bibliotheken in die Forschungsumgebung und das direkte Angebot einer Publikation (7) auf Edumeres.net erfolgt eine positive Bindung an das Portal. Des weiteren entsteht der Mehrwert einer einheitlichen Oberfläche für alle Stationen des wissenschaftlichen Forschungskreislaufes (Abbildung 3).



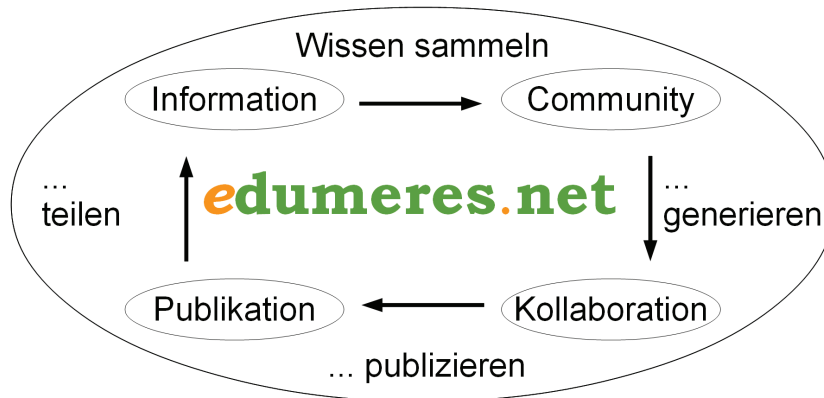


Abbildung 3: wissenschaftliche Wertschöpfung unter einer Oberfläche

Ebenfalls bereits anfänglich umgesetzt ist das Vorhaben der „Betreuten Projekte“. Das Web 2.0 ist zwar mittlerweile in vielen Bereichen Standard, dennoch liegt im wissenschaftlichen Bereich die Frustrationsgrenze in Bezug auf neue Formen der Kommunikation und Forschung niedrig. Aus diesem Grund ist es vorteilhaft jedem Projekt ein Mitglied der Redaktion zur Seite zu stellen, das von der Konzeption bis zur Publikation beratend tätig ist. Gleichzeitig ist es so möglich die VRE an die speziellen Erfordernisse eines jeden Forschungsvorhabens anzupassen, das heißt, die Arbeitswerkzeuge, soweit es im realisierbaren Rahmen liegt, projektspezifisch zu gestalten.

Dies führt zu den Planungen für die Zukunft. Gerade diese angesprochene Flexibilität wurde von den Nutzern begrüßt. Jedoch sind diese durch starre Insellösung Grenzen gesetzt. Deshalb wird bei der Weiterentwicklung darauf geachtet, dass das System Schnittstellen anbietet und nutzt, die eine Zusammenarbeit mit bereits vorhandenen Lösungen erlaubt. Als Beispiel sei hier nur der geplante Einsatz eines Sharepoint-Servers erwähnt, der in letzter Zeit weite Verbreitung fand und eine kostspielige Eigenentwicklung unnötig macht.

Was im Bereich der Technik angestrebt wird, ist auch auf institutioneller Ebene das Ziel. Zum einen gilt es die Synergieeffekte durch Kooperationen mit anderen Forschungseinrichtungen auszubauen und zu nutzen.

Zum anderen darf kein Mitglieder im Prozess der Wissensgenerierung, wie beispielsweise Bibliotheken und Infrastrukturdienstleister „nur“ als Quelle oder technische Voraussetzung gesehen werden. Erst durch deren konsequente Einbindung verdichtet sich das Netz zu einem tragfähigen Instrument der Forschungslandschaft. Erst so wird eine Virtuelle Forschungsumgebung, in diesem Falle Edumeres.net, zum Kommunikationsdispositiv mit dem Auftrag der Vernetzung von Daten, Menschen und Systemen.

### Literaturverzeichnis

- Allianzinitiative, Arbeitsgruppe 'Virtuelle Forschungsumgebungen' in der Schwerpunktinitiative 'Digitale Information'. "Definition Virtuelle Forschungsumgebung." *Allianz der deutschen Wissenschaftsorganisationen*. 2011. Zugriff: 15.08.2012.  
[http://www.allianzinitiative.de/de/handlungsfelder/virtuelle\\_forschungsumgebungen/definition/](http://www.allianzinitiative.de/de/handlungsfelder/virtuelle_forschungsumgebungen/definition/).
- Atkins, Daniel E. et al. "Revolutionizing Science and Engineering Through Cyberinfrastructure. Report of the National Science Foundation, Blue-Ribbon Advisory Panel on Cyberinfrastructure." *National Science Foundation* 2003. Zugriff: 08.08.2012.  
<http://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Carusi, Annamaria und Torsten Reimer. "Virtual Research Environment, Collaborative Landscape Study." *JISC*. 2010. Zugriff: 05.08.2012.  
<http://www.jisc.ac.uk/media/documents/publications/vrelandscape/vrelandscape.pdf>
- DINI, Arbeitsgruppe Virtuelle Forschungsumgebungen. "Virtuelle Forschungsumgebungen (vForum)". *Deutsche Initiative für Netzwerkinformation*. Zugriff: 07.08.2012. <http://www.dini.de/ag/vforum/>.
- Foucault, Michel, Gespräch mit Jacques-Alain Miller und anderen Vertretern des Teams Psychoanalyse der Universität Paris-VIII (1977). Abgedruckt in *Foucault-Handbuch, Leben - Werk - Wirkung* herausgegeben von Clemens Kammler, Rolf Parr und Ulrich Johannes Schneider, Stuttgart, 2008. 239.
- Fraser, Michael. "Virtual Research Environments: Overview and Activity." *Ariadne*. 2005. Zugriff: 05.08.2012. <http://www.ariadne.ac.uk/issue44/fraser>.
- GEI. "Das Institut." *Georg-Eckert-Institut*. Zugriff: 16.08.2012. <http://www.gei.de/das-institut.html>.
- Gneiting, Britta. Kollaboratives Wissensmanagement in einer Forschergruppe der Geisteswissenschaft. Magisterarbeit an der Universität Hildesheim, 2011.
- JISC. "Virtual research environment programme." *JISC*. Zugriff: 06.08.2012.  
<http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>
- Keller, Reiner, Andreas Hirsland, Werner Schneider und Willy Viehöfer. "Die diskursive Konstruktion von Wirklichkeit. Einleitende Bemerkungen zum Verhältnis von Wissenssoziologie und Diskursforschung." In *Die diskursive Konstruktion von Wirklichkeit. Zum Verhältnis von Wissenssoziologie und Diskursforschung*, herausgegeben von Dies. Konstanz, 2005. 7-21.
- Ruoff, Michael. *Foucault-Lexikon*. Paderborn, 2007.

- Vaart, Lilian van der, "Collaboratories: Connecting Researchers. How to facilitate choice, design and uptake of online research collaborations." *SURFfoundation*. 2010. Zugriff: 09.08.2012.  
<http://www.surf.nl/nl/publicaties/Documents/Collaboratories%20Connecting%20Researchers9april.pdf>
- Wolf, Werner. *Musicalization of Fiction. A Study in the Theory and History of Intermediality*. Amsterdam, 1999.
- Wolf, Werner. "Intermedialität. Ein weites Feld und eine Herausforderung für die Literaturwissenschaft." In *Literaturwissenschaft: intermedial - interdisziplinär* herausgegeben von Herbert Folzinek und Christoph Leitgeb. Wien, 2002. 163-192.

## **Fachspezifische Anpassungen einer pandisziplinären Forschungsplattform am Beispiel der Pathologie**

Mirosław Mainusch<sup>1</sup>, Dominik Schmitz<sup>1</sup>, Alberto Perez Bouza<sup>2</sup>, Saskia Eva Elisabeth Westphal<sup>3</sup>, Bernd Decker<sup>4</sup>

<sup>1</sup> RWTH Aachen University, Hochschulbibliothek, IT-Dezernat

<sup>2</sup> Universitätsklinikum Bonn, Institut für Pathologie

<sup>3</sup> RWTH Aachen University, Institut für Pathologie

<sup>4</sup> RWTH Aachen University, Rechen- und Kommunikationszentrum

### **Zusammenfassung**

Im vorliegenden Beitrag wird das DFG-Projekt "ProjektRepository" vorgestellt. Ziel ist die Entwicklung einer pandisziplinären Forschungsplattform auf Basis von MS SharePoint. Am Beispiel des Pilotpartners Pathologie wird erläutert, welche Möglichkeiten für fachspezifische Anpassungen angeboten und genutzt werden müssen, um den praxistauglichen Einsatz der Plattform in einer bestimmten Disziplin zu ermöglichen. Der Fokus liegt dabei auf dem Umgang mit (klinischen) Daten. Konkret wird die Plattform für die Kooperation von Pathologen und Radiologen sowie zu Lehr- und Ausbildungszwecken genutzt. Im Beitrag wird gezeigt, welche Rolle der standardisierten Erschließung im Hinblick auf eine einrichtungsübergreifende Interpretierbarkeit und Vernetzbarkeit von Forschungsdaten zukommt.

### **Abstract**

In this paper we present the DFG-funded project "ProjektRepository". The aim of the project is to develop a multidisciplinary research platform based on MS SharePoint. Focusing on one particular pilot partner, the institute of pathology, we explain how to configure and adapt the platform for a particular discipline in order to enable practical operational use. As the application field we focus on dealing with (clinical) data. The platform is used for cooperation between pathologists and radiologists, as well as for teaching purposes. The contribution emphasizes the important role of standardized descriptions particularly with regard to networking and interoperability of research data.

### **1. Einführung**

Virtuelle Forschungsumgebungen sind wichtige Instrumente der wissenschaftlichen Arbeit und sollten den Forschenden von ihren Einrichtungen bereitgestellt werden (Allianz 2011). Eine Universität steht dabei vor der Herausforderung, unterschiedliche Disziplinen mit ihren Spezifika angemessen bedienen und dies als verlässliche Leistung mit vertretbarem Aufwand anbieten zu müssen. Im Rahmen des DFG-Projekts „ProjektRepository“ erarbeiten das Rechen- und Kommunikationszentrum und die Bibliothek der RWTH Aachen das Grundmodell für eine

professionelle Forschungsumgebung, die leicht zu bedienen ist, weil sie das System der lokalen Lehr- und Lernplattform auf der Basis von MS Sharepoint nachnutzt und damit für die Pflege und Verwaltung Synergien schafft. Das Grundmodell hat einen pandisziplinären Ansatz und soll daher für alle an der RWTH vertretenen Fächer nutzbar sein. Exemplarisch wird es mit drei Forschergruppen aus unterschiedlichen Disziplinen entwickelt.

Als ein Partner in diesem Projekt ist das Institut für Pathologie in der Diagnostik, Forschung und Lehre tätig und organisiert interdisziplinäre Tumorkonferenzen. Solche interdisziplinären Forschungsprojekte, die sich mit Bildgebung in der Medizin beschäftigen, erfordern eine gemeinsame elektronische Plattform für die Auswertung der Ergebnisse, für den Vergleich einzelner Befunde und für die sowohl zeit- als auch raumunabhängige Diskussion.

Im Beitrag wird die prototypische Umsetzung einer solchen Plattform auf Basis des pandisziplinären Grundmodells dokumentiert. Gezeigt wird u.a., für welche Zwecke die Vernetzung von Systemen über Menschen bis hin zu Daten mithilfe der virtuellen Forschungsumgebung geeignet ist und welche fachspezifischen Features sich aus dem konkreten Anwendungsbeispiel ergeben. Wichtige Funktionen sind die Verwaltung gemeinsamer Daten, der wechselseitige Zugriff und die effektive Erschließung. Die Bibliothek sichert hierbei die nachhaltige formale und inhaltliche Erschließung und fungiert als Vermittler zwischen den fachspezifischen Anforderungen der einzelnen Disziplinen und dem pandisziplinären Ansatz des Rechen- und Kommunikationszentrums.

## **2. Verwandte Arbeiten in der Medizin (insbes. Pathologie)**

Berücksichtigt werden aktuelle Projekte, die sich vordergründig mit histologischen Mikroskopaufnahmen beschäftigen. Der Schwerpunkt der nachstehenden Darstellung liegt auf der Analyse der im Rahmen der Objektbeschreibung verwendeten Klassifikationen und Schemata (s. Tabelle 1).

**PathoPic** (1) ist ein deutschsprachiges Web-Portal mit einer öffentlich zugänglichen Bilder-Datenbank, die gegenwärtig 10.660 qualitativ hochwertige Mikroskopaufnahmen aus dem Bereich Pathologie umfasst. Die Bilder werden für den Einsatz im Bereich der medizinischen Ausbildung zur Verfügung gestellt. Jedes Bild der PathoPIC-Sammlung hat eine eindeutige Identifikationsnummer und wird

nach 10 Kategorien klassifiziert. Bei **Virtual Pathology at the University of Leeds** (2) handelt es sich um eine englischsprachige Online-Datenbank mit pathologischen Mikroskop-Aufnahmen. Sie umfasst gegenwärtig ca. 2.600 Aufnahmen. Jede Aufnahme wird anhand 8 Kategorien beschrieben. Ein einfach strukturiertes und schnell zu implementierendes Metadatenschema zur Beschreibung von Pathologie-Mikroskop-Bildern wurde von Jules J. **Berman** (3) vorgeschlagen. Um eine gute Beschreibung, Archivierung und Austauschbarkeit der Pathologie-Bilder zu ermöglichen, empfiehlt Berman die Verwendung von sieben fachspezifischen Kategorien. **PEIR** digital library (Pathology Education Instructional Resource) (4) enthält über 4.000 pathologische Bilder, die vor allem durch Angaben in Freitext-Form beschrieben werden. Jede Beschreibung enthält nur 3 Metadatenfelder. Bei **HEAL** (5) handelt es sich um eine Web-basierte Datenbank mit Multimediatdaten zur medizinischen Ausbildung. Die histologischen Schnitte werden von HEAL anhand sieben spezifischer Erweiterungen für mikroskopische Aufnahmen aus dem Bereich Pathologie beschrieben. Am Institut für Pathologie des Uniklinikums Aachen wurde die **Virtuelle Mikroskopie** (VM) (6) als Lehr- und Lerninstrument entwickelt. Die VM ist ein Online-Mikroskopie-Programm, in dem eingescannte histologische Schnitte wie unter einem echten Mikroskop betrachtet werden können. Dabei ist der Betrachter nicht auf einen vorgegebenen Bildausschnitt beschränkt und kann sich die Morphologie in verschiedenen Vergrößerungsstufen einprägen. Das ist wichtig, da häufig sowohl das Wachstumsmuster eines Tumors, als auch die Zellkernmorphologie für eine Diagnose von Bedeutung sind.

1. Pathopic	<a href="http://alf3.urz.unibas.ch/pathopic/intro.htm">http://alf3.urz.unibas.ch/pathopic/intro.htm</a>
2. VP Leeds	<a href="http://www.virtualpathology.leeds.ac.uk/">http://www.virtualpathology.leeds.ac.uk/</a>
3. Berman	<a href="http://www.julesberman.info/rdfimage.pdf">http://www.julesberman.info/rdfimage.pdf</a>
4. PEIR	<a href="http://www.peir.net/">http://www.peir.net/</a>
5. HEAL	<a href="http://www.healcentral.org/services/servicesMetadataSchema.jsp">http://www.healcentral.org/services/servicesMetadataSchema.jsp</a> . (zur Zeit nicht online)
6. VM RWTH	<a href="http://www.vm.rwth-aachen.de/intro/intro/index.html">http://www.vm.rwth-aachen.de/intro/intro/index.html</a>
7. Datacite	<a href="http://schema.datacite.org/">http://schema.datacite.org/</a>
8. Dublin Core	<a href="http://dublincore.org/">http://dublincore.org/</a>
9. LOM:	<a href="http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf">http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf</a>
10. Altman&King	<a href="http://www.dlib.org/dlib/march07/altman/03altman.html">http://www.dlib.org/dlib/march07/altman/03altman.html</a>
11. OECD	<a href="http://dx.doi.org/10.1787/787355886123">http://dx.doi.org/10.1787/787355886123</a>

Tabelle 1: Herangezogene Plattformen/Projekte im Bereich Pathologie (1-6) und fachübergreifende Metadatenschemata (7-11) [letzter Zugriff am 16.08.2012]

### **3. Ansatz**

Der im Projekt entwickelte Lösungsansatz besteht aus zwei zentralen Komponenten: Einer pandisziplinären Grundplattform auf Basis von SharePoint und der Möglichkeit zur Einrichtung von anwendungsbezogenen Forschungsräumen mithilfe fachspezifischer Konfiguration.

#### 3.1 Technologische Infrastruktur

SharePoint (Rizzo et al. 2010) ist ein weit verbreitetes Standard-Produkt für webbasierte Kommunikation und Kollaboration. Es bietet grundlegende Repository-Funktionen wie Verwaltung von Nutzern, Anlegen von Ordnerstrukturen und Objekten, Versionskontrolle, automatisierte Benachrichtigungsmöglichkeit und viele weitere, komplexe Funktionen. Als erste Anpassung für den Einsatz als virtuelle Forschungsumgebung wurde im Rahmen des Projekts das SharePoint-Framework um drei Versionen des Repository erweitert. Ziel ist es so variable Vernetzungsmöglichkeiten im Spektrum potentiell unterschiedlicher und sich im Projektverlauf ändernder Interessen der Forscher nach einem geschützten Raum anbieten zu können:

(I) Speicherndes Repository: Das speichernde Repository bietet die Möglichkeit Artefakte zu speichern und durch das Hinzufügen von Metadaten zu strukturieren.

(II) Vernetzendes Repository: Anders als das speichernde Repository wird im vernetzenden Repository nur eine Indirektion (Verweis/Link) auf das eigentliche Artefakt gespeichert.

(III) Repository Proxy: Der Proxy ermöglicht es, einen relevanten Auszug eines externen Repositories zu „ernten“, um darauf aufbauend lokal zu erschließen oder mit eigenen Repositories zu vernetzen. Sowohl die Artefakte als auch die Metadaten liegen beim Repository-Proxy extern vor.

Darüber hinaus müssen für die fachspezifische Konfiguration das Anlegen und Speichern unterschiedlicher, fachspezifischer Klassifizierungssysteme sowie die Festlegung geeignet angepasster Workflows für die Erschließung von Objekten ermöglicht werden.

#### 3.2 Fachspezifische Aspekte

Aufbauend auf dieser technologischen Grundlage zielen die Aufgaben der Bibliothek im Projekt auf eine präzise, fachspezifische Erschließung und die einrichtungsüber-

greifende Vernetzbarkeit von Objekten, hier konkret klinische Daten. Dies wird durch die Ermittlung und Anwendung möglichst standardisierter Metadaten erreicht. Als Metadaten werden Daten über Daten bezeichnet. Sie beziehen sich auf alle Informationen, die in strukturierter Form zur Beschreibung, Erläuterung, Verwaltung und Lokalisierung von Objekten verwendet werden (vgl. Greenberg; 2005; S. 20). Eine besondere Herausforderung bei der Beschreibung von Forschungsdaten stellt vor allem die Heterogenität dieser Daten zwischen den einzelnen Fachdisziplinen dar. Dies führt dazu, dass trotz einer Vielzahl von vorhandenen Standardisierungen eine Übertragung dieser in andere, neue Anwendungskontexte aufgrund besonderer Charakteristika der Forschungsdaten oft problematisch ist (vgl. Jensen; 2011; S.83). In bestimmten Fachdisziplinen wie z.B. den Sozial- und Geowissenschaften, liegen bereits fachübergreifende und gut etablierte Standardisierungen zur Beschreibung von (Forschungs-)Daten vor. Etwa kann auf Schemata wie Data Documentation Initiative (DDI) bzw. ISO 19115 oder INSPIRE-Richtlinie zurückgegriffen werden. Sind solche fachspezifischen Schemata in einer Disziplin nicht vorhanden, müssen sie unter Berücksichtigung fachspezifischer Kontexte in einem Auswahl- und Beschreibungsprozess für die jeweiligen Forschungsdaten entwickelt werden (vgl. TIB et al. 2010). Zur Unterstützung der Interoperabilität der zu definierenden Schemata sollte dabei die Entwicklung beispielsweise entlang den Zertifizierungskriterien des Kompetenzzentrum Interoperable Metadaten (KIM) erfolgen, die wiederum auf dem Interoperabilitätsmodell Singapore Framework basieren (KIM 2010).

#### 4. Fachliche Erschließung medizinischer Daten am Beispiel Pathologie

Der Erfassungsprozess von medizinischen Informationen ermöglicht prinzipiell synonyme Bezeichnungen, abweichende Notationsweisen sowie gewisse Spielräume bzgl. Umfang und Struktur der Beschreibungen. Die damit verbundene Redundanz erschwert die maschinelle Verarbeitung der Daten und ein präzises Wiederfinden einzelner Objekte bzw. Befunde. Ein typisches Beispiel ist ein bösartiger Tumor des Nasen-Rachen Raumes, das sog. nasopharyngeale Carcinom (syn. Schmincke-Tumor, lymphoepitheliales Carcinom, Lymphoepitheliom). Dieser Tumor wird nach ICD-O je nach morphologischer Variante mit den Nummern 8072/3, 8071/3 oder 8083/3 verschlüsselt.



Des Weiteren ergeben sich bei einer nicht standardisierten Beschreibung Einschränkungen bei der Durchführung von Häufigkeitsanalysen, da bei der wissenschaftlichen Auswertung die Gleichwertigkeit bestimmter Merkmalsbeschreibungen ggf. nicht gewährleistet werden kann. Steht der Informationsaustausch zwischen mehreren, an der Behandlung oder Forschung beteiligten Einrichtungen im Vordergrund, kann die Implementierung unterschiedlicher Bezeichnungen zudem zu gravierenden Verständigungsschwierigkeiten und Missverständnissen führen (vgl. Leiner 2006; S. 30f.)

#### 4.1 Ziele und Vorgehen

Die in der Diskussion verwandter Arbeiten (s. Kap. 2) angesprochenen Projekte aus dem Bereich Pathologie zeigen, dass der Einsatz von computergestützten Methoden in der Medizin eine Vielzahl neuer Möglichkeiten etwa zur diagnostischen Auswertung oder in der Lehre eröffnet. Zugleich erfordert dies aber auch eine sorgfältige fachspezifische Objektbeschreibung. Die medizinische Objektbeschreibung muss dabei den jeweiligen Verwendungszweck ausreichend berücksichtigen. Daher müssen bei der Umsetzung im vorliegenden Projekt zunächst die angestrebten Ziele präzise identifiziert werden.

(I) Als primäres Ziel wurde festgelegt, dass makroskopische oder histologische Merkmale eines bestimmten Organs (z.B. Herz oder Niere) mit der radiologischen Bildgebung direkt verglichen werden können, um eine radiologisch-pathologische Befundkonkordanz oder -diskrepanz festzustellen.

(II) Des Weiteren wurde festgelegt, dass die auf der Plattform gesammelten Objekte samt ihrer Beschreibungen neben der Nutzung in der klinisch-wissenschaftlichen Forschung auch zur klinischen Aus- und Fortbildung eingesetzt werden sollen.

In beiden Szenarien kommt dem Datenaustausch eine wesentliche Bedeutung zu. Daher wurde angestrebt mithilfe standardisierter Beschreibung eine möglichst breite Interoperabilität und Vernetzbarkeit zu erreichen. Dies kann dabei vor allem durch Standardisierung auf zwei Ebenen erreicht werden: auf der formalen und der inhaltlichen Ebene.

#### 4.2 Formale Beschreibung: Definition/Auswahl von Metadatenfeldern

Eine wesentliche Rolle spielen dabei Metadatenschemata. Ein Metadatenschema kann beschrieben werden als „[...] eine systematische, geordnete Kombination von

Elementen oder Begriffen“ (Woodley et al., 2005) und „ein Satz von Metadatenelementen sowie Regeln für Ihre Nutzung“ (NISO, 2004, S. 16). Um ein für das Projekt geeignetes Schema zu definieren wurde im ersten Schritt analysiert, welche Metadatenfelder (Merkmale) mit welchen Ausprägungen in Bezug auf die in dem Projektkontext gegebenen Objekte zu berücksichtigen sind.

Da bei makroskopischen und histologischen Bildern eine sinnvolle Interpretation erst durch zusätzliche Informationen möglich ist, die das jeweilige Bild in seiner Entstehung sowie den abgebildeten Inhalt beschreiben, ist es wichtig, einen geeigneten Kernsatz von Metadatenfeldern für makro- oder mikroskopische Aufnahmen aus dem Bereich Pathologie zu identifizieren. Dieser muss alle relevanten Inhalte eines Bildes präzise beschreiben, darf zugleich aber nicht zu komplex sein, um im praktischen Einsatz unmittelbar Anwendung finden zu können. Auch Angaben wie Alter und Geschlecht des vorliegenden Patienten sind für die Diagnosestellung sehr wichtig.

Die Recherche und nachfolgende Analyse der Metadatenschemata im Bereich Pathologie führten zum Ergebnis, dass gegenwärtig im gegebenen Verwendungskontext kein geeignetes Metadatenchema für histologische Mikroskopaufnahmen existiert. Um die für eine Beschreibung der Histologie-Aufnahmen notwendigen Metadatenfelder zu definieren, wurden zunächst die fachübergreifenden Metadatenschemata (s. Tabelle 1) von DataCite (7), DublinCore (8) -sowie das LOM-Metadatenchema (9) analysiert. Berücksichtigt wurden zudem Metadatenchema nach Altman und King (10) sowie das OECD-Metadatenchema (11). Für die Entwicklung einer Metadatenbasis im Bereich medizinischer und technischer Metadaten für pathologische Mikroskopaufnahmen wurden die bereits im Kontext verwandter Arbeiten vorgestellten Plattformen (s. Tabelle 1, (1)-(6)) herangezogen. Auch wenn diese Beschreibungen keinem etablierten Schema-Standard folgen, liefert ihre Analyse dennoch erste Orientierungspunkte für die zu definierenden Metadatenfelder. Aus dieser Analyse entstand ein erster Metadatenschema-Entwurf, der ca. 45 Metadatenfelder umfasste. Nach einer Validierung dieses Schemas durch die Fachwissenschaftler aus dem Bereich Pathologie erfolgte eine bedarfsorientierte Anpassung, im Rahmen derer mehrere weitere Metadatenfelder, vor allem aus dem Diagnostikbereich ergänzt, technische Metadaten dagegen in Unterstrukturen zusammengefasst wurden. Aktuell umfasst das Metadatenschema 26 Felder (s. Tabelle 2).

Medizinische Metadatenfelder	Beschreibung / Beispiel	Kontrolliertes Vokabular
Krankheitsgruppe (ICD10)	Neubildungen, Bestimmte infektiöse und parasitäre Krankheiten, Störungen des Immunsystems, etc.	19 Werte
Fach	Medizinische Fachrichtung, z.B. Pathologie	3 Werte
Topographiegruppe	Topographische Regionen, z.B. Haut, Kardiovaskuläres System	18 Werten
Betrachtung Organsystem	Methode der Darstellung oder Zugangsweg, z.B. Endoskopie, Operation.	10 Werte
Diagnose	Befundinterpretation, z.B. Myokardinfarkt, Lungenembolie	179 Werte
Lokalisation/ Tumorklassifikation des Primärtumors (ICD-O-3)	Genaue Zuordnung zu einer anatomischen Lokalisation, z.B. C18 Dickdarm allgemein, C18.0 bis C18.8 Dickdarm in verschiedenen Regionen unterteilt	130
Tumormorphologie (ICD-O-3)	Interpretation von Wachstumsmuster und Erscheinungsbild der Tumorzellen, z.B. 8070/3 für Plattenepithelkarzinom, 8070/2 für Carcinoma in situ der Plattenepithels	163 Werte
BI-RADS	Breast Imaging Reporting and Data System z.B. BI-RADS 2 für gutartigen Befund	7 Werte
B-Klassifikation	Klassifikation für die histologische Diagnose von Mamma-stanzbiopsien des NHSBSP, z.B. B2 für gutartigen Befund	5 Werte
Seite	Seitenangabe der Brust, z.B. rechts	2 Werte
Alter	Alter des Patienten. Einordnung in eine vordefinierte Altersgruppen	8 Werte
Geschlecht	Geschlecht des Patienten, z.B. weiblich	3 Werte
Färbung	Die Art der Schnittfärbung, bei der ein Farbstoff, der in Lösung angeboten wird, an definierte Gewebestrukturen bindet	27 Werte
Immunhisto	Anfärbung von Zellbestandteilen mit Antikörpern	113 Werte
Verschlagwortung nach MESH (DC.Subject)	Verschlagwortung (nach Bedarf) mithilfe des polyhierarchischen, konzeptbasierten Thesaurus MESH	57.656 deutsche Entry Terms 181.141 englische Entry Terms
Vergrößerung der Histologie	Histologische Präparate können in hoher Auflösung untersucht werden. Das Feld definiert die vorhandene Vergrößerung	7 Werte
<b>10 weitere organisatorische und techn. Metadatenfelder</b>	z.B. Titel, Ersteller / Urheber, Herausgeber, Rechte, Ressourcentyp.	<b>Freitext, URL</b>

Tabelle 2: Entwickeltes Metadatenschema Pathologie

16 davon entfallen auf die medizinische Beschreibungsebene, wohingegen 10 Felder die Bilder innerhalb der Sammlung spezifizieren und Auskunft über administrative Zusammenhänge geben. Ähnlich wie in den Datenbanken Virtual Pathology und PEIR sollen die technischen Metadaten etwa zu Bildformat, Bildgröße, Vergrößerung, Mikroskop, Filter, Kamera, Objektiv, Scanner-Gerät, relevanter Scanbereich etc. möglichst automatisch ermittelt und in die Beschreibung auf der Forschungsplattform übernommen werden.

Durch die auf diese Weise entwickelte Beschreibung mit Metadaten können histologische Bilder nun in verschiedene Kontexte gebracht werden. Es wird dabei zwischen der patientenbezogenen und der patientenübergreifenden Fragestellung unterschieden. Bei der patientenbezogenen Fragestellung werden das histologische Bild und dessen mögliche Differenzialdiagnosen in einen Kontext mit den zuvor gestellten Diagnosen des Patienten gebracht (z.B. handelt es sich um einen neu entstandenen Tumor oder um eine Absiedlung eines zuvor bei diesem Patienten diagnostizierten Tumors?). Hierzu sind Metadaten wie Lokalisation, Seite, Tumormorphologie, Immunhistologie, Alter und Geschlecht des Patienten relevant. Bei der patientenübergreifenden Fragestellung wird das histologische Bild z.B. mit den möglichen Differenzialdiagnosen in einen Kontext gebracht (gleichet das histologische Bild einem Tumor, der bei einem anderen Patienten diagnostiziert wurde, oder in einem Lehrbuch abgebildet ist?). Dabei sind die medizinischen Metadaten wie Fach, Betrachtung, Tumormorphologie, Färbung, Vergrößerung, Immunhistologie, BI-RADS und B-Klassifikation wichtig (s. Kap. 4.3).

#### 4.3 Inhaltliche Beschreibung: Klassifikationen und Thesauri

Im Rahmen der vorgestellten Analyse wurde festgestellt, dass sowohl der Verwendung medizinischer Klassifikationen wie auch der präzisen Verschlagwortung histologischer Mikroskopaufnahmen essentielle Bedeutung für das richtige Verständnis und die Interpretierbarkeit dieser Bilder zukommt. Nur die Verwendung von Begriffen aus kontrollierten Vokabularen, Terminologiesystemen und verbreiteten Klassifikationen kann die semantische Standardisierung konkreter Merkmalsausprägungen und folglich ihre Vergleichbarkeit und Interoperabilität garantieren (vgl. Schöch 2008, S. 14).

Die nachstehenden, medizinischen Klassifikationen wurden für den Einsatz im Rahmen des Projekts ausgewählt:

(I) Die ICD-O (International Classification of Diseases for Oncology 3rd Edition) als eine duale Klassifikation, da sowohl Schlüssel für die Lokalisation als auch Schlüssel für die Histologie einer Neubildung enthalten sind (DIMDI 2003). Die ICD-O wird international zur Verschlüsselung von Krebsregistern und Krebsmortalität eingesetzt. Jeder pathologische Befund eines malignen Tumors muss die entsprechende ICD-O Kodierung für Tumorentität und Tumorlokalisierung enthalten.

(II) Der zusätzliche Einsatz des einachsigen und monohierarchischen Klassifikationssystems ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) ermöglicht eine Verknüpfung der einzelnen Bilder mit dem Gegenstandskatalog (GK), der die Prüfungsinhalte der Staatsexamina im Studiengang Medizin regelt (DIMDI 2011).

(III) Zur Verschlagwortung der Bilder werden Medical Subject Heading (MeSH)-Codes eingesetzt (NLM 2011). Dies ermöglicht eine Anbindung an die zentrale medizinische Publikationsdatenbank PubMed. So kann etwa der Befund: Dieulafoy's malformation mit Hilfe einer vorformulierten Anfrage mit den diagnoserelevanten Beständen von PubMed verlinkt werden.

(IV) Die BI-RADS Klassifikation (Breast Imaging Reporting and Data System) ist eine Klassifikation des American College of Radiology für die Befundung von Mammographien durch Radiologen (American College of Radiology 2003). Sie soll durch eine einheitliche Terminologie des Mammographie-Befundes die Kommunikation zwischen Ärzten vereinfachen und standardisierte Behandlungsschemata ermöglichen.

(V) Die internationale B-Klassifikation ist eine fünfstufige Klassifikation für die histologische Diagnose von Mammastanzbiopsien (NHSBSP 2001). Die B-Klassifikation ersetzt nicht die endgültige histopathologische Diagnose, sondern reduziert den Befund auf eine der fünf Kategorien, so dass der Pathologe gezwungen ist, sich auf eine Kategorie festzulegen. Die Klassifikation ist wichtig für weitere therapeutische oder diagnostische Maßnahmen und standardisierte Behandlungsschemata.

(VI) Als zukünftige Erweiterung ist die Einbindung von UMLS angedacht (NCBI 2009). Das Unified Medical Language System (UMLS) ist ein Metathesaurus, der Notationen aus etwa 100 begrifflichen Ordnungssystemen, Thesauri und Klassifikationen verbindet. Dies ermöglicht eine effiziente Verknüpfung von Benutzeranfragen und Fachinformationen sowie Überführung vorhandener Beschreibungen in andere Begriffssysteme.

#### 5. Erste Erfahrungen

Im Laufe des Projekts hat sich die Erstellung der Klassifikationen und Metadaten-schemata in SharePoint als sehr komplex erwiesen. Daher wurde SharePoint um eine Import- und Export-Schnittstelle für die Web-Ontology-Language (OWL) (W3C 2009) erweitert. Durch die Schnittstelle ist es möglich die notwendigen Schemata in dafür vorgesehenen Ontologie-Editoren wie z.B. Protégé zu erstellen und diese dann als Struktur für das Repository zu nutzen. Zusätzlich ist es durch den Export möglich, das Metadaten-schema und das Repository samt Inhalt weiter zu verwenden, etwa zur Vernetzung mit anderen Repositories oder um die Inhalte für z.B. semantische Suchmaschinen erschließbar zu machen.

Da die umfangreichen und großen Bilddaten (mehrere GB pro Einzelbild möglich) bereits in geeigneter Umgebung im Uniklinikum Aachen abgelegt sind, wurde für den Datenaustausch die Version vernetzendes Repository gewählt. So werden die bereits durch die klinische Routine vorhandenen Daten und Bilddateien mit Metadaten angereichert und weiteren ärztlichen Kollegen zur Verfügung gestellt. Zum Anwendungsgebiet gehören neben der Forschung auch interdisziplinäre Tumorkonferenzen. Hier besprechen Fachärzte verschiedener Disziplinen (z.B. niedergelassene oder universitäre Chirurgen, Onkologen, Strahlentherapeuten, Radiologen und Pathologen) die Befunde eines Patienten, um gemeinsam eine Empfehlung für die optimale Therapie zu erarbeiten. Dabei werden histologische und radiologische Bildgebung verglichen und Befundkonkordanz oder -diskrepanz festgestellt. Mithilfe einer vernetzten Datenbank können sich alle an der Konferenz beteiligten Ärzte optimal in den Fall einarbeiten, was bislang aufgrund von Zugriffsbeschränkungen auf die unterschiedlichen Krankenhaus- oder Praxisinformationssysteme nicht möglich ist. Hier zeigt sich aber zugleich auch eine Einschränkung der Nutzbarkeit des ProjektRepository für die klinische Verwendung.

Aus Gründen des Datenschutzes dürfen Patientendaten nicht unter Verwendung personenbezogener Daten in einem Netzwerk freigegeben werden. Die Gefahr, dass die Identitäten von Patienten öffentlich werden, muss mit völliger Sicherheit ausgeschlossen werden können. Dagegen ist es unproblematisch, pseudonymisierte Daten in der Forschung oder für eine sekundäre Verwendung in der Lehre oder Weiterbildung zu nutzen. In diesem Zusammenhang ist geplant, eine „Trainingsplattform“ für Ärzte in Weiterbildung und Studenten aufzubauen. Mithilfe dieser Plattform sollen diese sich mit dem Thema der interdisziplinären Tumorkonferenzen auseinandersetzen und eine eigene Einschätzung der zu einem Fall verfügbaren Bilddaten abgeben können.

#### 6. Zusammenfassung

Der vorliegende Beitrag zeigt am Beispiel der Medizin, dass die fachspezifischen Anforderungen an eine Forschungsplattform insbesondere im Hinblick auf die Beschreibung von Forschungsdaten besonders berücksichtigt werden müssen, um die Akzeptanz der Wissenschaftler zu erhalten. Das im Projekt entwickelte Grundmodell wurde daher um entsprechende Möglichkeiten zur fachspezifischen Konfiguration erweitert. Eine zentrale Rolle bei der Beschreibung spielen die fachspezifischen Metadatenschemata, Klassifikationen, Thesauri und kontrollierten Vokabulare, die die Vergleichbarkeit und die gemeinsame Nutzung der Daten gewährleisten. Das Beispiel histologischer Mikroskopaufnahmen zeigt zudem, dass das Fehlen von etablierten Metadatenschemata, deren Strukturen die besonderen Charakteristika der jeweiligen Forschungsdaten ausreichend berücksichtigen, in bestimmten Disziplinen nach wie vor eine nicht zu übersehende Herausforderung darstellt. Die Konzeption eines dem Nutzungsszenario angemessenen Schemas erfordert zum Teil aufwändige Recherche- und Analyse-Prozesse sowie anschließende Eigenentwicklungen. Entsprechend muss die Plattform geeignete Schnittstellen bereitstellen, um der Bibliothek bzw. dem Wissenschaftler, auch wenn sie nicht über das technische Know-How verfügen, die Möglichkeit zu geben, die konzipierten Beschreibungsmöglichkeiten in die Plattform zu integrieren.

## Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen, 2011. Online:  
[http://www.allianzinitiative.de/fileadmin/user\\_upload/2011\\_VRE\\_Definition.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/2011_VRE_Definition.pdf) [Zugriff am 16.08.2012]
- American College of Radiology, 2003. ACR BI-RADS®- Mammography: Breast Imaging Reporting and Data System, Breast Imaging Atlas, vol 4. ACR, Preston.
- Greenberg, Jane, 2005. Understanding Metadata and Metadata Schemes. In: Smiraglia, Richard (Hrsg.), 2005: Metadata: A Cataloguers's Primer; Haworth Information Press, Binghamton
- NHSBSP, 2001. Guidelines for Non-operative Diagnostic Procedure and Reporting in Breast Cancer Screening. Online: <http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp50.pdf> [Zugriff am 08.08.2012]
- DIMDI, 2003. ICD-O 3 Auflage. Online:  
<http://www.dimdi.de/dynamic/de/klasi/downloadcenter/icd-o-3/1auflage2003/> [Zugriff am 16.08.2012]
- DIMDI, 2011. ICD- 10 Version 2011. Online:  
<http://www.dimdi.de/dynamic/de/klasi/downloadcenter/icd-10-who/version2011/systematik/x1wbp2011.zip> [Zugriff am 16.08.2012]
- Jensen, U.; Katsanidou, A.; Zenk-Möltgen, W.: Metadaten und Standards. In: Büttner, S.; Hobohm, H.; Müller, L., 2011. Handbuch Forschungsdatenmanagement; Bock+Herchen, Bad Honnef.
- KIM, 2010. Kriterien für die Zertifizierung von Metadatenprofilen. Online:  
[http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Berichte/kriterienFuerDieZertifizierungVonMetadatenprofilen.pdf?\\_\\_blob=publicationFile](http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Berichte/kriterienFuerDieZertifizierungVonMetadatenprofilen.pdf?__blob=publicationFile) [Zugriff am 01.08.2012]
- Leiner, Florian, 2006. Medizinische Dokumentation. Schattauer, Stuttgart
- National Center for Biotechnology Information (NCBI), 2009. UMLS® Reference Manual. Online: <http://www.ncbi.nlm.nih.gov/books/NBK9684/> [Zugriff am 10.08.2012]
- National Information Standards Organization (NISO), 2004. Understanding Metadata. Online:  
<http://www.niso.org/publications/press/UnderstandingMetadata.pdf> [Zugriff am 10.08.2012]
- NLM, 2011. Online <http://www.nlm.nih.gov/mesh/filelist.html> [Zugriff am 15.08.2012]
- Rizzo, T.; Alirezai, R.; Fried, J.; Swider, P.; Hillier, S.; Schaefer, K., 2010. Professional Sharepoint 2010 Development. Wrox.
- Schöch, Winfried, 2008. Einsatz von DICOM-SR in der Pathologie. Online:  
[http://www.schoech.de/diploma/Schoech\\_2008\\_Using\\_DICOM\\_SR\\_in\\_Pathology.pdf](http://www.schoech.de/diploma/Schoech_2008_Using_DICOM_SR_in_Pathology.pdf) [Zugriff am 01.08.2012]
- Technische Informationsbibliothek (TIB), FIZ Chemie, Universität Paderborn, 2010. Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie. Online:  
[http://www.tib-hannover.de/fileadmin/projekte/primaer-chemie/Konzeptstudie\\_Forschungsdaten\\_Chemie.pdf](http://www.tib-hannover.de/fileadmin/projekte/primaer-chemie/Konzeptstudie_Forschungsdaten_Chemie.pdf) [Zugriff am 01.08.2012]
- W3C, 2009. OWL 2 Web Ontology Language. Online:  
<http://www.w3.org/TR/owl2-overview/> [Zugriff am 16.08.2012]
- Woodley, M; Clement, G.; Winn, P., 2005. DCMI-Glossary. Online: <http://dublincore.org/documents/2005/11/07/usageguide/glossary.shtml> [Zugriff am 01.08.2012]



---

## **Vom „embedded“ zum „liaison librarian“ – was versprechen die neuen Konzepte?**

Simone Fühles-Ubach

Fachhochschule Köln, Institut für Informationswissenschaft

### **Zusammenfassung**

In den vergangenen Jahren wurde über den „embedded“ librarian diskutiert und auch darüber, dass die Einbindung der Bibliothekare in Projekt-, Ausschuss- oder Fakultätsarbeit direkt erfolgen muss, um die Bibliothek bei Wissens- und Informationsfragen der Trägerinstitution adäquat einzubinden. Seit 2009 propagiert die ARL ein weitergehendes Konzept, den sog. „liaison librarian“. Ein Sonderheft berichtet ausführlich, wie sich insbesondere im Bereich der Forschungs- und Universitätsbibliotheken eine weitere starke Veränderung des Tätigkeitsprofils der Bibliothekare hin zu sogenannten „Liaison Librarians“ entwickeln und verändern wird. „Building relationships is becoming the essence .. - one that connects users with their information needs, ..“

Das neue Konzept zielt darauf, dass Bibliothekare individuelle Kundenbindungen aufbauen und besonders die „elektronische“ Bibliothek auf diese Bedürfnisse zuschneiden. In der Universitätsbibliothek von Christchurch (Neuseeland) wurde dieses Konzept radikal umgesetzt, nachdem zwei Erdbeben Teile der Bibliothek stark beschädigt und unbenutzbar hinterlassen haben. Die Liaison Librarians kontaktieren jeden Professor mindestens einmal, bei Interesse auch zweimal pro Jahr, erheben den Informationsbedarf mittels einer spezifischen Interviewtechnik, beraten aktiv zum Bestand und besprechen Neuanschaffungen, die vorwiegend im Bereich der elektronischen Medien liegen sollen. Bereits nach weniger als einem Jahr findet das Konzept größte Zustimmung der Kunden.

Der Beitrag thematisiert die Ablösung des Konzepts des „embedded“ hin zum „liaison librarian“ und dessen Tätigkeitsprofilen. Darüber hinaus wird erörtert, welche Rahmenbedingungen und Gründe für die Einführung dieses Konzeptes sprechen könnten und wie neu dieses Vorgehen tatsächlich ist. Im Mittelpunkt des Konzeptes steht hier der ausnahmsweise mal „weakest link“: die elektronischen Bestände, die ohne aktives Marketing „unsichtbar“ als Bibliotheksservice bleiben.

### **Abstract**

In the last years the concept of „embedded librarians“ was much discussed and it was stated that a straight links of librarians to projects, committees and faculty work is a prerequisite for the integration of librarians in knowledge and information businesses of their responsible bodies. Since 2009 ARL has announced a secondary concept called „liaison librarian“. A special issue provided detailed information about the changes for librarians especially in the university and research libraries. „Building relationships is becoming the essence .. - one that connects users with their information needs, ..“

The new concept aims at building individual relationships between librarians and customers and at creating electronic libraries according to customer needs. The university library of Christchurch, New Zealand, put this concept into practice after two massive earth quakes left the library badly damaged and no longer usable. The liaison librarians contact every professor at least once, in case of greater interest even twice a year, they gather the information needs via special interviews, give advice about the collection and talk about new acquisitions, which are mainly electronic media. Even after less than one year the concept meets the general approval of the customers.

The article discusses the change of the concepts from „embedded“ to „liaison librarian“ and its fields of activity. Furthermore the overall conditions and reasons for introducing this concept are presented. The main aspect is the „weakest“ link, the electronic collections, which remain invisible as library services without active marketing

### **Ausgangspunkt**

Die Art unserer gesamten Kommunikation und unser Informationsverhalten verändern sich. Das ist für Bibliothekare keine große Neuigkeit, aber dennoch hilft es hin und wieder, sich die Geschwindigkeit der Veränderung noch einmal zu vergegenwärtigen. Erst dann wird die Herausforderung klar, die es bedeutet, „am Ball“ oder „auf der Höhe der Zeit“ zu bleiben. Vor wenigen Wochen feierte Apples iPhone seinen fünften Geburtstag und damit wurde die Ära der Smartphones eingeläutet. Jeder dritte Bundesbürger besitzt bereits ein Smartphone, in der Gruppe der unter 30-jährigen, d.h. dem großen Teil der primären Zielgruppe von wissenschaftlichen Bibliotheken, sind es bereits mehr als die Hälfte (Tuma, 2012, S.63). Schätzungen gehen davon aus, dass es in ca. 5 Jahren keine anderen Handys mehr geben wird.

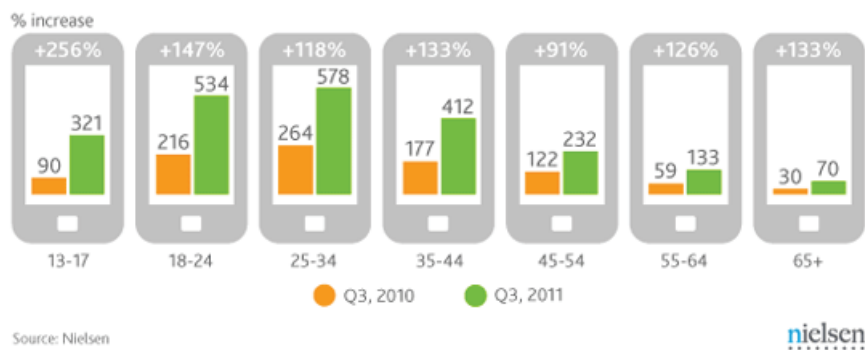
Dabei ist weniger die Tatsache, dass sich etwas verändert problematisch, als vielmehr die Geschwindigkeit der Veränderung. Während der Buchdruck als revolutionäre Veränderung noch zwei Jahrhunderte benötigte, um weltweit Verbreitung zu finden, benötigte der Hörfunk nur noch zwanzig Jahre. (ebd.) Beim World Wide Web, dessen „Geburtstag“ man auf den 30.04.1993 datiert, weil es an diesem Tag erstmals für die Allgemeinheit zur Verfügung gestellt wurde, waren es weniger als 10 Jahre und – das Smartphone kann nun mit nur fünf Jahren als eine der schnellsten Innovationen gelten.

Informationssuche findet daher bereits in vielen Fällen – und das wird sich aufgrund dieser Entwicklung noch verstärken – quasi in der Hosentasche statt und die Bibliotheken sind gehalten, wenn nicht gar gezwungen, sich auf diese Neuerung einzustellen.

Nicht genug mit den technischen Innovationen. Auch die Benutzer und ihr Kommunikations- und Informationsverhalten ändern sich. Laut Nielsen (2011) tauschen Teenager in den USA (13-17 Jahre) **stündlich** im Durchschnitt 7 SMS oder MMS aus und das bei einer Tageslänge von 16 Stunden pro Tag. Gegenüber den Vorjahreswerten (3. Quartal 2010) ist dies eine Steigerung um 256% und diese Generation drängt ab dem kommenden Jahr an die Hochschulen.

#### Monthly Data Usage by Age (MB)

Q3, 2010 vs. Q3, 2011



Wie die Grafik zeigt, sind die Steigerungsraten nicht nur in dieser Altersgruppe erheblich: Bei fast allen Altersgruppen gibt es eine Steigerung um mehr als 100% , nur die 45-55jährigen liegen mit 91% Steigerung knapp darunter.

Wie das Informationsverhalten der 18-24jährigen aussieht, hat eine von der IMLS Rutgers University durchgeführte Studie herausgefunden (Radford, 2010). Die 18-24jährigen, d.h. diejenigen sich gerade an den Universitäten befinden, wurden nach ihrer Bewertung der virtuellen Auskunftsdienste befragt. Dabei waren sowohl Nutzer als auch Nicht-Nutzer in die Untersuchung einbezogen. Methodisch basiert die Studie auf Fokusgruppen-Interviews, der Analyse von Transkripten sowie Online-Befragungen und Telefoninterviews. Diese Gruppe wurde zwischen 1988-1994 geboren und kann sich nicht an eine Zeit ohne Computer erinnern.

Die Fokusgruppen-Interviews ergaben folgendes Suchverhalten:

- sie bevorzugen Google & Suchmaschinen,
- stöbern im Netz,
- fragen Freunde,
- finden ihre Informationen selbst.

Hinsichtlich der Bibliotheksnutzung ergab sich ein ernüchterndes Bild: Obwohl der größte Teil der Befragten Handys besaß, war ihnen nicht bekannt, dass es einen telefonischen Auskunftsdienst gibt – niemand hat jemals in der Bibliothek für Hilfe bei den Hausaufgaben (Übungen) angerufen oder wäre auf die Idee gekommen, eine Mail an die Bibliothek zu schreiben. Darüber hinaus gab es interessanterweise Bedenken, dass die Bibliothekare sie nicht verstehen würden oder nicht ernst nehmen würden (Face to Face ist einschüchternd) oder es wurden (Daten-) Sicherheitsbedenken genannt.

Die Online-Befragung ergab, dass der Chat von den Nutzern als beste der virtuellen Auskunftsdienste eingeschätzt wird. Dabei wurde der virtuelle Auskunftsdienst meist deshalb genutzt, weil er von jemand konkret empfohlen wurde. Problematisch fanden sie insbesondere nicht hilfreiche Antworten (Informationen schon selbst gefunden), wenn die Antwortenden sich nicht als Spezialisten herausstellten oder wenn die Verbindung zu langsam war. Nicht-Nutzer gaben an, dass sie nichts von der Existenz dieses Dienstes wussten, sich nicht sicher sind, ob Bibliothekare helfen können, kein 24/7-Service angeboten wird und sie mit ihren eigenen Ergebnissen zufrieden sind. Die Frage: „Was würde den virtuellen Auskunftsdienst attraktiv oder noch attraktiver machen, bringt folgende Ergebnisse“:

- schnellere und einfachere Software
- personalisierte Schnittstellen
- längere Servicezeiten
- erfahrene, freundliche und technik-affine Bibliothekare

Dezentralisierung der Bibliothek als Rahmenbedingung

Was bedeuten all diese Entwicklungen und Einschätzungen für das Management von Bibliotheken, d.h. für die Steuerung der Prozesse und die Führung des Personals? Die Entwicklung in den verschiedenen Bibliothekstypen verläuft hier sehr unterschiedlich.

2009 hat eine Studie zur Zukunft von Forschungsbibliotheken ergeben, dass sich die Bibliotheken dort besonders schnell und stark wandeln (Fühles-Ubach / Lorenz, 2009, 9). Die untersuchten Beispiele (u.a. Bank, Schwerindustrie, Pharmakonzern) zeigten, dass die Forschungsbibliothek der Zukunft keine „vier Wände“ und vor allem keinen Lesesaal mehr haben wird, in dem sich die Kunden vor Ort informieren. Damit verschwinden auch die Kunden aus der Bibliothek, da die angebotenen Dienstleistungen über das Netz (Internet / Intranet) oft im benötigten Volltext im eigenen Büro bzw. Forschungsumfeld verfügbar gemacht werden. Die Bibliothek als konkreter Ort spielt für die Forscher kaum noch eine Rolle. Sie wünschen sich und brauchen die „library at your desk“. Die Dienstleistungen werden zwar noch zentral beschafft und organisiert, jedoch nicht mehr zentral in der Bibliothek, sondern über das Netz dezentral direkt beim Kunden vor Ort angeboten. Als konkreten Beleg wird man die Schließung der Fraunhofer Zentralbibliothek in St. Augustin im Frühjahr diesen Jahres betrachten müssen.

Deutlich anders stellt sich die Situation in den Hochschulbibliotheken dar. Zwar sind auch hier die elektronischen Medien und die virtuellen Besuche(r) stark im Wachstum begriffen und bestimmte Dienstleistungen wie Katalogrecherchen werden zunehmend von außerhalb der Bibliothek wahrgenommen. Dennoch verzeichnen die Lesesäle aber eine zunehmende Frequentierung und die Bibliothek als Arbeits- und Aufenthaltsort hält nicht nur, sondern steigert sogar noch ihre Bedeutung.

In einer repräsentativen Studie für die Ruhr-Universität Bochum (Fühles-Ubach, 2011) wurden die Studierenden nach der Nutzung der Lern- und Leseplätze befragt, d.h. nach der Bibliothek als Lernumgebung und Aufenthaltsort. Im Ergebnis blieben die meisten Befragten bis zu drei Stunden pro Bibliotheksbesuch im Lern- und Lesebereich der Universitätsbibliothek (Median bei drei Stunden, Doktoranden 3,5 Stunden). Darüber hinaus zeigen sich interessante Entwicklungen innerhalb des Studienverlaufs. In den ersten Semestern des Studiums werden die Lehr- und Lernplätze vor Ort signifikant häufiger genutzt als in späteren Semestern. Die Gründe hierfür könnten darin liegen, dass die Studierenden zu Beginn des Studiums noch stark darauf angewiesen sind, einen Überblick über das eigene Studienfach und seine Teilgebiete zu entwickeln.

Hierbei helfen die umfangreichen Lehrbuchsammlungen und die großen Printbestände, die in den Regalen einen Überblick über vorhandene Literaturschwerpunkte sichtbar machen (*Browsing*).

Mit zunehmender Studiendauer wird der Überblick größer, und es entwickeln sich Schwerpunkte und spezifische Forschungsfragen. Als Folge davon wird auch der Literaturbedarf spezieller, so erhalten z.B. die Fachzeitschriften einen höheren Stellenwert für die eigene Arbeit. Da diese häufig online vorliegen, werden die Studierenden und die Wissenschaftler in zunehmendem Maße unabhängig vom Bibliotheksort. Innerhalb der Bibliothek benötigen sie jedoch eine fachlich stärker fokussierte und spezialisierte Beratung vor Ort. Hochschulbibliotheken werden sich daher zukünftig immer stärker zu einer Kombination aus virtuellem und physischem Bibliotheksraum mit Fachberatung entwickeln (ebd. S.11f).

Damit entwickeln sich deutlich zwei verschiedene Schwerpunkte:

- Steigerung / Ausbau digitaler Dienstleistungen um Zugriff von außen zu gewährleisten
- Ausbau der Bibliothek als Lernort mit kompetenter Fachberatung, um das Lernen innerhalb des Gebäudes zu optimieren

Ein ganz extremes Beispiel für die Neuorientierung hin zu digitalen Dienstleistungen bietet die Universitätsbibliothek von Christchurch; Neuseeland. Dort wurde das Bibliotheksgebäude durch zwei Erdbeben im Februar und Juni 2011 stark beschädigt, der Benutzerverkehr musste phasenweise ganz eingestellt werden und ist noch heute nur eingeschränkt möglich, das 10stöckige Gebäude, das große Teile des Bestandes enthält, kann nur vom Bibliothekspersonal mit besonderen Vorsichtsmaßnahmen betreten werden. Um den Studienbetrieb dennoch aufrechterhalten zu können und für die Lehrenden und Studierenden eine angemessene Literaturversorgung zu schaffen, musste der größtmögliche Teil der Medien in ein digitales Umfeld überführt werden und diese Plattform musste den Lehrenden und Studierenden in kürzester Zeit nahe gebracht werden. Um dies zu erreichen, wurde das Konzept des „liaison librarian“ in weniger als 6 Monaten mit großem Erfolg umgesetzt.

Doch was ist ein liaison librarian und wie unterscheidet sich das Konzept vom „embedded librarian“, das noch vor einem Jahr als aktuell und zukunftssträftig vorgestellt wurde? (Salz, 2011)

### **Stärkere Einbindung in Projekte und im Unternehmen - der „embedded librarian“**

Die Veränderung der Bibliothek als Ort führt dazu, dass Bibliothekare ihre Leistungen nicht mehr vorwiegend innerhalb der Bibliothek erbringen, sondern sich näher am Arbeitsbereich der Kunden befinden und sich in Projekten oder größeren Besprechungen mit ihren Angeboten und Dienstleistungen aktiv beteiligen und einbringen. Dieser Status der Einbindung in Projekt- und Unternehmensstrukturen wird als „embedded librarian“ bezeichnet. In der konkreten Umsetzung konnte diese Situation bereits bei einem Pharma-Unternehmen, einer Bank und einer Unternehmensberatung beobachtet werden. In allen Institutionen ist die Bibliothek / Informationsvermittlungsstelle zwingend bei der Eröffnung von Projektvorhaben zu beteiligen und darüber hinaus werden regelmäßig Präsentationen über die neuen oder sich verändernden Dienstleistungen angeboten (halbjährlich). (Fühles-Ubach / Lorenz, 2008, S.9)

Über die Umsetzung am Forschungszentrum Jülich wurde bei der letzten AspB-Tagung ausführlich berichtet. Hier wurde ein „embedded librarian“ als Informationsmanager im Vorstandsbüro eingesetzt, der alle bibliothekarischen bzw. informationsbezogenen Tätigkeiten unmittelbar in die allgemeinen Arbeits- und Projektprozesse des Kunden integriert und unmittelbar vor Ort erbringt. Auf diese Weise werden die dortigen Mitarbeiter von allen Tätigkeiten rund um die Informationsbeschaffung, -verarbeitung und – nachbereitung entlastet. (Salz, 2011, S. 3)

Dabei wurde in Jülich ein zentraler Gesichtspunkt berücksichtigt, indem der Informationsbedarf des Vorstandsbüros und damit der Top-Management-Ebene besondere Berücksichtigung gefunden hat. Auf diese Weise wird bibliothekarische Kompetenz und Erfahrung aus der Zentralbibliothek in den Vorstandsbereich transponiert und gleichzeitig eine dauerhafte Anbindung und Einbindung in die Prozesse der Zentralbibliothek gewährleistet.



#### Ausgangspunkt Strategie

Das Konzept des „liaison librarian“ geht über dieses Konzept des „embedded librarian“ deutlich hinaus. Es geht dabei weniger um die Integration von einzelnen Personen in den weiteren Unternehmenskontext, sondern vielmehr um eine Veränderung der Arbeits- und Kommunikationsweise von großen Teilen der Bibliotheksmitarbeiter. Vergleichbar ist dieses Konzept vielleicht mit dem eines militärischen „Verbindungsoffiziers“, dessen Aufgabe u.a. das Halten und Pflegen von dienstlichen Kontakten zu anderen Teilbereichen der eigenen Institution ist. Er ist besonders wichtig bei gemeinsamen „Operationen“ und bringt die Interessen und Angebote ihrer entsendenden Einrichtung mit ein.<sup>1</sup>

Die guten „Beziehungen“ sind so wichtig, weil der Erfolg einer Informations- und Forschungseinrichtung maßgeblich von ihrem Kontakt zur Trägereinrichtung abhängt und vor allem davon, in welcher Weise sie zur Erreichung der strategischen Zielplanung der Institution beitragen kann. Diese Verbindung verdeutlicht die Grafik.



---

<sup>1</sup> Verbindungsoffizier Zivil-Militärische Zusammenarbeit. -  
<http://www.reservisten.bundeswehr.de/portal/a/resarb/resoffz/aktuelle/zivilmil/verbindu1>  
(letztmalig recherchiert am 13.08.12)

Hervorzuheben ist die starke Betonung des Bezugs zum Unterhaltsträger, der in erster Linie durch den Nachweis der Mittelverwendung ein Interesse an der Leistungsdarstellung für die Gesamtorganisation und für die primäre Nutzergruppe besitzt.

### Die neue Rolle von Bibliothekaren

Wie die Rolle des „liaison librarian“ für Bibliotheken definiert werden kann, zeigt die kurze Aufgabenbeschreibung von Kara M. Whatley, Bibliothekarin und Leiterin des Coles Science Center der New York University:

*“Building relationships is becoming the essence of what it is to be a liaison librarian—one that connects users with their information needs, whatever the format and whatever the technology.” (Whatley, 2009, 29 f.)*

Der amerikanische Berufsverband der Forschungsbibliotheken, ARL – Association of Research Libraries“ hat im Sommer 2009 ein Sonderheft herausgebracht, indem ausführlich darüber berichtet wird, wie sich insbesondere im Bereich der Forschungs- und Universitätsbibliotheken eine starke Veränderung des Tätigkeitsprofils der Bibliothekare hin zu sogenannten „Liaison Librarians“ entwickeln und verändern wird. Karla Hahn, Assistant Executive Director für den Bereich “Research, Teaching, and Learning” in der ARL zeichnet folgendes Szenario:

*... several new roles that are being fostered in research libraries and offer different leadership perspectives on the change process.....*

*...New kinds of relationships are needed to respond to the changing work of faculty and researchers and to constantly evolving learning outcomes, research processes, and communication practices. In addition, research libraries are increasingly seeking to influence larger dialogs about changing practices that are occurring on campuses and within disciplines. **Liaison librarians need well-developed, high-trust relationships to create strategic opportunities to participate in and influence disciplinary and departmental decisions.** It is also evident that the range of activities addressed through liaison work is growing. As a consequence, liaison librarians increasingly need the ability to acquire new skills and leverage more specialized expertise among their library colleagues in service of their clients. Liaisons cannot be expert themselves in each new capability, but knowing when to call in a colleague, or how to describe appropriate expert capabilities to faculty, will be key to the new liaison role. Just as researchers are often working in teams to leverage compatible expertise, liaison librarians will need to be team builders among library experts where this advances client research. New liaison roles are not emerging de novo, but rather in continuity with established roles. (Hahn, 2009, S.1f)*

Damit wird nach dem "embedded librarian", dessen interne Vernetzung in der Organisation und vor allem in der Projektarbeit stark betont wurde, eine weitere Komponente in das Berufsbild integriert: Die Rolle des Beziehungsarbeiters in der eigenen Organisation und auch in spezifischen Außenbereichen.

In einem Positionspapier, dem Position Description Framework werden 10 Aufgabenbereiche für „Liaison Librarians“ definiert (Williams, 2009, 3-8):

1. Campus Engagement:  
Es wird mehr Engagement in Gremien und Bereichen außerhalb der Bibliothek gefordert, um näher an der Trägerinstitution und deren Gesamtausrichtung zu sein.
2. Collection Development and Management  
In Absprache mit den Benutzern werden die Bestände entwickelt (s. Patron Driven Acquisition) und (hinsichtlich Medienform und Standort) verwaltet.
3. Teaching and Learning  
Den Benutzer innerhalb der Bibliothek für die eigenständige Arbeit mit den Bibliotheksservices, wie z.B. Recherche in internationale Fach-Datenbanken, fit zu machen.
4. Scholarly Communication  
Stärkere und engere Kundenbindung zu Forschern (und Studierenden).
5. Digital Tools  
Erkennen von Neuigkeiten auf dem Markt und deren Nutzungspotentialen und Anpassungsoptionen für die eigenen Einrichtung
6. Reference/Help Services  
Elektronische Auskunftsdienste für inhaltliche und technische Fragen zu den Informationsquellen.
7. Outreach (to the local community)  
Öffentlichkeitsarbeit (innerhalb der eigenen „Gemeinschaft – hier ist durchaus auch die Trägerorganisation in engeren Sinne sowie die Kommune im weiteren Sinne gemeint)
8. Fund Raising  
Einwerben von Spenden und Geldern zur finanziellen Förderung
9. Exhibition and Event Planung  
Die Bibliothek soll sowohl Eigeninitiative bei Ausstellungsarbeit zeigen, als auch Beteiligung an Veranstaltungen der Gesamtorganisation zeigen.
10. Leadership  
Übernahme von Führungsaufgaben nicht nur innerhalb, sondern auch außerhalb der Bibliothek in Gremien sowie Vorreiterrolle bei allen Aspekten die Medien- und Informationskompetenz der Gesamtorganisation.

### **Die Reduktion bestehender Tätigkeitsfelder**

Mit der Übernahme neuer Aufgaben geht in der Regel auch eine Reduktion bisheriger Tätigkeiten einher. Dazu wurden die statistischen Auswertungen der ARL aus den vergangenen 17 Jahren herangezogen, die zeigen, welche Tätigkeiten sich im Verlauf der vergangenen Jahre verstärkt oder reduziert haben.

In der Kategorie „Service“ hat sich der Leihverkehr zwischen den Bibliotheken stark erhöht. Ebenso die Arbeit für Gruppenpräsentationen, d.h. Schulungen. Gleichzeitig steigt die Anzahl der Studierenden, die die Dienste nutzen. Der Umschlag der Medien hat leicht zugenommen. Gleich geblieben ist die Anzahl der Mitarbeiter in diesem Bereich. Stark zurückgegangen sind die Auskunftsdienste, d.h. der Bereich der als „Reference Service“ subsummiert wird. (ARL, 2009, 9)

Diese Entwicklung bestätigt den aufgezeigten Trend der stärkeren (Kunden)-Beziehungsarbeit in Form von Präsentationen und Schulungen, während das klassische „Auskunftsgeschäft“ aufgrund der eigenen Internet-Recherchen der Benutzer stark zurückgeht. Was bleibt sind weniger, aber qualitativ deutlich anspruchsvollere Recherchen.

Bei den Ausgaben sind die Kosten für Zeitschriften sehr stark gestiegen, ebenso für „Bibliotheksmaterial“ (z.B. Netzwerktechnik), aber auch die gesamten Betriebskosten. Gehälter und Monographien konnten ebenfalls in diesem Zeitraum einen sehr leichten Anstieg verzeichnen. (ebd., S.15)

Betrachtet man die Ausgaben pro Student, so wurden deutlich weniger Monographien gekauft und es reduzierte sich das Verhältnis von Student / Forscher zum Bibliothekspersonal. Pro 1000 Besucher stehen 11 Mitarbeiter zur Verfügung (ebd., S. 16).

Besonders genau werden die Ausgaben für elektronische Ressourcen betrachtet, da sie seit 2008 mehr als die Hälfte der Ausgaben der Bibliothek ausmachen. Mit 51% aller Ausgaben stellen sie den größten Etatposten der Bibliothek dar. Unterschieden werden die elektronischen Ressourcen nach ebooks, eserials, bibliothekarisches Material (Netzwerk intern / extern), Hard- und Software, Document Delivery / ILL. (ebd., S.18)

Zusammenfassend lässt sich die Situation in den amerikanischen Bibliotheken wie folgt beschreiben: Die Kunden nutzen die Bibliotheksdienste anders als bisher. Es besteht ein wesentlich stärkerer Austausch zwischen den Bibliotheken zur Deckung der Kundenbedarfe.

Die Einzelauskünfte an Benutzer gehen hingegen zurück. Im Bestand dominieren weiterhin die Kosten für Zeitschriften, während der Kauf von Monographien und damit auch deren relativer Anteil am Erwerbungsetat zurück geht.

Mehr als die Hälfte des Etats wird mittlerweile für elektronische Ressourcen ausgegeben, die jedoch für die Leistungsmessung schwierig zu erfassen sind. Zum einen sind sie für Kunden und Träger als Bestand **nicht sichtbar** und verlangen deshalb eine ausgeprägte **Kommunikation über Vorhandensein und Nutzung**. Zum anderen sind sie ortsunabhängig in der Nutzung, d.h. sie werden **überwiegend außerhalb der Bibliothek** genutzt, die damit physische Kunden verliert und nur wenig Feedback über die Durchführung und den Erfolg der Informationsvermittlung hat.

Besonders wichtig ist dabei, dass 64% aller befragten Kunden der amerikanischen Bibliotheken die elektronischen Ressourcen außerhalb des Bibliotheksgebäudes nutzten. Weiter differenziert nach Nutzergruppen lässt sich feststellen, dass studentische Nutzung im undergraduate Bereich noch am stärksten innerhalb der Bibliothek zu finden ist, während graduate students and researchers die elektronischen Ressourcen nicht nur von außerhalb der Bibliothek, sondern auch von außerhalb des Universitätsgeländes nutzen.(Brinley, 2006, 49) Das entspricht den bereits zitierten Ergebnissen der Studie der Ruhr-Universität Bochum.

### **Konsequenzen für die Bibliotheksarbeit**

Welche Veränderungen bringt eine Verwirklichung des „liaison librarian-Konzeptes“ mit sich? Voraussetzung ist zunächst eine nüchterne Betrachtung der Fakten und Veränderungen im Bibliotheksumfeld und die Bereitschaft, zugunsten neuer Entwicklungen „alte“ Dienstleistungen zu reduzieren oder sogar einzustellen, um Raum für persönliche Kommunikation zu schaffen.

Aus den verschiedenen Trends wird ja deutlich, dass nicht alle Dienstleistungen hauptsächlich virtuell gewünscht sind. Ein Beispiel außerhalb des Bibliotheksbereiches sind z.B. Shopping-Malls, die sich zunehmender Beliebtheit erfreuen, auch wenn es für alle Artikel die Option gibt, sie online über das Internet einzukaufen. Was im Netz fehlt ist allerdings die persönliche Beratung und die Option des „Vor-Ort-Anschauens“ oder Probierens.

Genau diesen Vorteil bietet auch der „Lernort und Aufenthaltsort Bibliothek“, der den Studierenden eine ideale Lern- und Forschungsumgebung mit Fachberatung im adäquaten Umfeld bietet. Dieser Aspekt betrifft die Kommunikation mit den Studierenden.

Zusätzlich verändert sich aber auch die Kommunikation mit den Fach- und Finanzverantwortlichen, wie z.B. Lehrstuhlinhabern, Institutsleitern oder Dekanen. Hier geht es um den Aufbau kontinuierlicher Betreuungsrelationen mit regelmäßigen Treffen, um Wünsche und Bedarfe abzuklären. Der Trend zur „Patron Driven Acquisition“ ist hier bereits ein richtiger Weg. Wie in einer „liaison“ geht es um die „Zeit zu zweit“, in der die Bibliothek die Möglichkeiten hat, Fachgespräche zu führen und auf diese Weise z.B. Dienstleistungen oder Produkte, die der Kunde nicht sehen kann, bekannt zu machen oder zu bewerben, die der Kunde nicht sehen kann, wie z.B. neue elektronische Services oder Bestände. Den Spezialkenntnissen der Fachreferenten kommt hier wieder eine größere Rolle zu, um mit Wissenschaftlern und Forschern auf Augenhöhe kommunizieren zu können.

Für die Bibliotheksleitung und auch das mittlere Management der Bibliothek bedeutet dies einen Paradigmenwechsel hinsichtlich der Kommunikation nach außen. Kunden werden – ähnlich wie bei Versicherungen oder anderen Dienstleistungen, die man nicht sehen kann – im Einzelverhältnis mehr oder weniger intensiv und regelmäßig betreut, um so langfristige und stabile „Beziehungen (=liaisons) zu bauen. Dabei sind der „embedded“ und der „liaison“ librarian Konzepte, die nicht per se für eine Institution richtig oder falsch sein können. Vielmehr ergeben Sie sich aus dem jeweiligen Umfeld und seinen Optionen und Rahmenbedingungen. Nicht überall ist es möglich, wirkliche „Beziehungsrelationen“ aufzubauen. Hier verhält es sich wie im richtigen Leben – Beziehungen funktionieren nur dann gut, beide Parteien sich auf ein gemeinsames Konzept einigen können – und davon kann keineswegs immer und überall ausgegangen werden.

## Literaturverzeichnis

ARL Statistics 2007-2008, compiled and edited by Martha Kyrillidou, Les Blend, - Association of Research Libraries, 2009. - 152 S. – S. 9 <http://www.arl.org/bm~doc/law08.pdf>

Brinley Franklin and Martha Kyrillidou. "Demonstrating Value and Creating Value: Evidence-Based Library Management through MINES for Libraries™", Presented at the Scholars Portal Forum, Ontario Council of University Libraries, Toronto, Canada, February 1, 2006. – 65 slides. P. 49ff [http://www.libqual.org/documents/admin/OCULMINESPresentation\\_Final.ppt](http://www.libqual.org/documents/admin/OCULMINESPresentation_Final.ppt)

Fühles-Ubach, Simone; Lorenz, Miriam: „Die Forschungsbibliothek der Zukunft oder – Es ist eine ungeheure Stärke, wenn man nicht Recht haben muss. - Vortrag im Rahmen des 98. Bibliothekartages in Erfurt am 05.06.2009; Tagungsband 2009. Vortragsfolien unter: <http://www.opus-bayern.de/bib-info/volltexte/2009/798/>

Fühles-Ubach, Simone: Online-Kundebefragung Universitätsbibliothek Bochum. 2011. <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/projects/benutzererumfrage2011.pdf>. (letztmalig recherchiert am 08.08.2012)

Hahn, Karla: Introduction: Positioning Liaison Librarians for the 21st Century. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 265 (August 2009): 1–2. <http://www.arl.org/resources/pubs/rli/archive/rli265.shtml>

Nielsen: (2011) [http://blog.nielsen.com/nielsenwire/online\\_mobile/new-mobile-obsession-u-s-teens-triple-data-usage/](http://blog.nielsen.com/nielsenwire/online_mobile/new-mobile-obsession-u-s-teens-triple-data-usage/) (letztmalig recherchiert am 08.08.2012)

Radford, Marie L.; Lynn Silipigni Connaway. (2010): Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-User, and Librarian Perspectives. - <http://www.oclc.org/research/activities/synchronicity/ppt/lida2010.pptx> (letztmalig recherchiert am 08.08.2012)

Salz, Edith: "Embedded Librarian im Forschungszentrum: Ein Praxisbericht" in Spezialbibliotheken -- Freund und Follower der Wissenschaft : Präsentationen der 33. Arbeits- und Fortbildungstagung der Arbeitsgemeinschaft der Spezialbibliotheken e.V. -- Sektion 5 im Deutschen Bibliotheksverband / JARA - Forschungszentrum Jülich und RWTH Aachen University (Ed.) Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag : 2011 ; ASpB 2011: Spezialbibliotheken - Freund und Follower der Wissenschaft, 09.-11. November 2011, Auditorium der Zentralbibliothek im Forschungszentrum Jülich ; 9 <http://hdl.handle.net/2128/4491> (letztmalig recherchiert am 08.08.2012)

Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-User, and Librarian Perspectives. – Project Website: <http://www.oclc.org/research/projects/synchronicity> (letztmalig recherchiert am 08.08.2012)

Tuma, Thomas (2012): iPhone, also bin ich.  
In: Der Spiegel. – 27 / 2012. – S. 62-72.

Whatley, Kara M.: New Roles of Liaison Librarians: A Liaison's Perspective. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 265 (August 2009): 29–32. <http://www.arl.org/resources/pubs/rli/archive/rli265.shtml>. (letztmalig recherchiert am 08.08.2012)

Williams, Karen: A Framework for Articulating New Library Roles. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 265 (August 2009): 3–8. <http://www.arl.org/resources/pubs/rli/archive/rli265.shtml> (letztmalig recherchiert am 08.08.2012). O N

## 24/7/365 x 360 – die neue Zauberformel?

Anne-Katharina Weilenmann

BIBLINK.ch

### Zusammenfassung

Multitasking, fortgesetzte partielle Aufmerksamkeit oder doch lieber „Power Browsing“?

Die „Digital Natives“ bewegen sich in einer Lebens- und Gedankenwelt, die geprägt ist durch Unmittelbarkeit, Schnelligkeit, Informations- und Reizüberflutung. Sie fordern viel von ihrem Umfeld, werden aber auch stark gefordert; soziale Netzwerke wollen gepflegt sein, permanente Interaktion mit Gleichaltrigen ist wichtigstes Gebot, es ist ein Pendeln zwischen der virtuellen und realen Welt, in der die Grenzen fließend sind.

Die „Millennials“ sind es längst gewohnt, mit all diesen Einflüssen umzugehen und sich den Gepflogenheiten ihrer Generation anzupassen. Diverse Studien belegen, dass diese jungen Leute ein anderes Verhalten zeigen als ältere Jahrgänge („Digital Immigrants“), dass sie eigene Strategien entwickeln bei der Benutzung neuer Technologien und der Bewältigung der Informationsflut. Intuitive, einfache Zugangswege zu den Informationskanälen werden verlangt, Qualitätsbeurteilung rückt in den Hintergrund, man gibt sich schnell zufrieden mit dem Gefundenen. Dreh- und Angelpunkt für Recherchen sind kaum die Bibliotheken, sondern Suchmaschinen.

Sind wissenschaftliche Bibliotheken und Informationseinrichtungen obsolet geworden, vermögen sie den Strom fortwährender pulsierender Kommunikation der jungen Generation nicht mehr aufzunehmen? 24/7/365 x 360, 24 Stunden am Tag, 7 Tage die Woche, 365 Tage im Jahr, 360°-Panorama, die Erreichbarkeit über alle Kanäle, ist dies wohl die Formel der Zukunft, die nahtlose und dauerhafte Interaktion mit den Bibliothekskunden? Braucht es nicht auch wieder vermehrt die Verbindung zum Menschen, die Kontakte zum Einzelnen, die in einer entpersonalisierten Umgebung neu aktiviert und belebt werden wollen? Der persönliche Informationsberater, der „personal coach“, und massgeschneiderte, auf die individuellen Bedürfnisse abgestimmte Dienstleistungspakete werden zunehmend an Bedeutung gewinnen.

Im Beitrag werden Möglichkeiten erkundet, als Informationseinrichtung diesen Herausforderungen zu begegnen und optimale Angebote zu entwickeln.

### Abstract

Multitasking, continuous partial attention or power browsing?

Digital Natives know the restless life, ubiquitous proximity, immediacy is the important thing, they have to cope with too much information and attractions. They are always on, blogging, twittering, chatting in social networks; it's for them like a permanent shift from the virtual to the real world.



Millennials show unique search strategies, they demand user-friendly retrieval systems. Relevance of search results is less important, students are highly satisfied with the results. They don't visit libraries, the first destination for answers is the search engine. Are academic libraries out-of-date, are their services too old-fashioned for the Net Generation? 24/7/365 x 360, permanent access and support on all devices, can this be the new magic formula for the library to reach their patrons? Or is there a need for a „personal information coach“? Individual support and personalized solutions for patrons will be the great challenge. The article discusses possibilities to meet these challenges.

### **Einleitung**

Signale, Bits, Datenströme beherrschen zunehmend unser Leben, prägen unseren Alltag. Dabei ist jedes Individuum selbst Teil dieser Datenströme, versucht sich einzuklinken in den steten Fluss, schwimmt mit diesen Strömen, um keine Welle zu verpassen. Permanente Erreichbarkeit scheint das höchste Gut zu sein, 24/7/365 x 360, 24 Stunden am Tag, 7 Tage die Woche, 365 Tage im Jahr, 360°-Panorama, Kommunikation über alle Kanäle (Kühne 2011).

Dieses Bild zeigt sich insbesondere bei der jungen Generation, die in der „Mindset List 2016“<sup>1</sup> wie folgt beschrieben wird: „They have always lived in cyberspace, addicted to a new generation of ‘electronic narcotics’.“<sup>2</sup> Ungestüm, ihrem Umfeld alles abverlangend, ein pulsierendes Lebensgefühl; nichts scheint unmöglich zu sein in der Welt der Internetgeneration... Was heisst das für Bibliotheken und Informationseinrichtungen, wie können sie die Anzeichen dieses Wandels aufnehmen, was können sie einer rastlosen ungeduldigen Bevölkerungsschicht bieten?

### Digital Natives, Digital Visitors, Digital Residents...

Geboren im Zeitalter von Internet, Informations- und Kommunikationstechnologien, aufgewachsen mit Handy und entsprechenden Gadgets, zeigen die Jungen von heute ein gänzlich neues Verhalten im Umgang mit Medien. Prensky (2001) nennt diese Gruppe „Digital Natives“, der digitalen (Mutter)Sprache von Computer, Videospielen und Internet mächtig, und unterscheidet sie von den „Digital Immigrants“, denjenigen, welche

---

<sup>1</sup> Die „Mindset List“ wird jährlich seit 1998 vom Beloit College erstellt, mit dem Ziel, die Weltsicht und Ideen der neu eintretenden Studierenden zu erfassen. Die aktuelle Ausgabe 2016 berücksichtigt den Jahrgang der 1994 Geborenen.

<sup>2</sup> S. Beloit College 2012, o.S.

die digitale Welt erst spät(er) in ihrem Leben kennenlernen und erkunden können. Er beschreibt die „Digital Natives“ als „... accustomed to the twitch-speed, multitasking, random-access, graphics-first, active, connected, fun, fantasy, quick-payoff world of their video games, MTV, and Internet are bored by most of today's education, well meaning as it may be.“<sup>3</sup>

Die Terminologie Prenskys wird jedoch zunehmend von kritischen Stimmen hinterfragt und aufgelöst. White/Le Cornu (2011) prägen die Begriffe „Digital Visitor“ und „Digital Resident“. Für die beiden Autoren ist nicht die Altersfrage das entscheidende Element bei der Differenzierung, sondern Kontext und Motivation bei der Benützung technischer Hilfsmittel. So macht sich der „Digital Visitor“ die Neuen Informations- und Kommunikationstechnologien zu Nutze, wenn er sie benötigt (Funktion als Werkzeug, Tool): „... Visitors understand the Web as akin to an untidy garden tool shed. They have defined a goal or task and go into the shed to select an appropriate tool which they use to attain their goal. Task over, the tool is returned to the shed.“<sup>4</sup> Während dessen sieht der „Digital Resident“ eine Verweilplattform darin (space, place): „Residents, on the other hand, see the Web as a place, perhaps like a park or a building in which there are clusters of friends and colleagues whom they can approach and with whom they can share information about their life and work.“<sup>5</sup> Die unterschiedlichen Verhaltensweisen dieser zwei Gruppen können aufschlussreiche Erkenntnisse liefern und sind Gegenstand weiterer Untersuchungen (White et al. 2012).

Die immer grössere Vielfalt an technischen Möglichkeiten verleitet dazu, die ganze Bandbreite an Diensten und Anwendungen voll auszureizen; am kreativsten funktioniert dies, wenn alles parallel und nebeneinander zum Einsatz kommt. Diese simultane Geschäftigkeit, auch als „Multitasking“<sup>6</sup> bezeichnet, bedeutet nicht nur eine Effizienz- und Produktivitätssteigerung unserer Aktivitäten, sondern wir verfolgen für die einzelnen Tätigkeiten ein konkretes Ziel und messen ihnen die gleiche Priorität bei (Small/Vorgan 2009). Davon klar abzugrenzen ist die fortgesetzte partielle Aufmerksamkeit.

---

<sup>3</sup> S. Prensky 2001, S. 5.

<sup>4</sup> S. White/Le Cornu 2011, o.S.

<sup>5</sup> S. ibid., o.S.

<sup>6</sup> „Multitasking“ wird nach „Wissen.de“ wie folgt definiert: 1. <EDV> gleichzeitiger Betrieb mehrerer, voneinander unabhängiger Programme, Aufgaben; 2. <allg.> das Ausführen mehrerer Befehle oder Tätigkeiten zum gleichen Zeitpunkt. Eine der wenigen Untersuchungen zum Thema Multitasking wird 2010 unter dem Titel „Generation M<sup>2</sup>“ von Rideout et al. (2010) durchgeführt.

Im Gegensatz zum Multitasking schenken wir nach Stone (2005)<sup>7</sup> dabei allem und jedem dauerhaft unsere Aufmerksamkeit, können uns trotzdem auf nichts wirklich festlegen und konzentrieren: „To pay continuous partial attention is to pay partial attention — CONTINUOUSLY. It is motivated by a desire to be a LIVE node on the network. Another way of saying this is that we want to connect and be connected.“<sup>8</sup> Wir sind nach allen Seiten vernetzt, hängen in diesem Netz, verheddern uns zuweilen darin. Neben der Bewältigung der üblichen Aufgaben kann dies zu einer Herausforderung, schnell auch zu einer Überforderung führen: „It is an always-on, anywhere, anytime, any place behavior that involves an artificial sense of constant crisis.“<sup>9</sup>

Die permanente Benützung der sozialen Medien kann gerade auch im Studienalltag leicht zuviel werden; dies belegen u.a. Aussagen wie „The wrong, untamed use of multi-tasking is becoming an epidemic,...“<sup>10</sup> Weiter findet dieselbe Studentin: „Educators ‘should start offering a class or coaching in middle/high schools on how to effectively manage multi-tasking, how to organize and discipline yourself for blocks of uninterrupted time in their work day.’ This is, she added, ‘fast becoming a skill the new generation will need in order to manage their success and well-being in the adult world’.“<sup>11</sup>

Die Frage stellt sich, wie Universitäten und andere Bildungsinstitutionen die aktuellen und zukünftigen Strömungen und Trends im Bereich der Neuen Informations- und Kommunikationstechnologien aufnehmen und in die Curricula integrieren, welche Auswirkungen dies in weiterer Konsequenz auf die Bibliotheken hat, und insbesondere welche neuen Herausforderungen und Chancen damit verbunden sind.

„Given the range and volume of information available and the ubiquity of access to information sources and resources, learning strategy shifts from a focus on information as such to judgment concerning reliable information, from memorizing information to how to find reliable sources. In short, learning is shifting from learning that to learning

---

<sup>7</sup> Der Begriff „fortgesetzte partielle Aufmerksamkeit“ (englischer Originalausdruck: continuous partial attention) wird von der ehemaligen IBM-Managerin Linda Stone geprägt; weitere Ausführungen dazu sind in ihrem Blog nachzulesen: <http://lindastone.net/> (Stand 1. September 2012).

<sup>8</sup> S. Stone 2005, o.S.

<sup>9</sup> S. ibid., o.S.

<sup>10</sup> S. Brokaw 2012, o.S.

<sup>11</sup> S. ibid., o.S.

how, from content to process.“<sup>12</sup> Die Verschiebung weg von inhaltsbasiertem zu prozessorientiertem Denken sowie der zunehmende Einsatz von E-Learning-Plattformen und Online-Tools bringt eine Veränderung der Lernkultur mit sich. Die Bildungslandschaft ist einem grossen Wandel unterworfen, die Universitäten einem immer stärker werdenden Konkurrenzdruck ausgesetzt. Begünstigt wird diese Entwicklung durch Anbieter von Lerninhalten wie Udacity<sup>13</sup> oder Coursera<sup>14</sup>, die sich auf qualitativ hochwertige Kurse spezialisiert haben und mit Slogans wie „the World's Best Courses, Online, For Free“<sup>15</sup> werben. Die Dozenten<sup>16</sup> sind mit Überlegungen konfrontiert, ob und mit welchem Nutzen neue Kommunikationsformen wie Wikis, Blogs, Twitter usw. in die Lehre eingebracht werden können. Zurzeit ist diesbezüglich noch eine grosse Zurückhaltung festzustellen, kaum jemand bringt den Mut auf, hier einen oder auch mehrere Schritte in die Zukunft zu wagen; man kann höchstens von einzelnen experimentellen Versuchen sprechen.<sup>17</sup>

Laut Einschätzungen von Experten wird sich die Tätigkeit des Lehrens und Lernens bis zum Jahre 2020 stark wandeln, im Mittelpunkt steht der Student der Internetgeneration mit seinen individuellen Bedürfnissen und Anliegen. Prägende Faktoren dieser Umwälzungen werden sein: individualisiertes, personalisiertes Lernen (der Student stellt sich seine Lerninhalte nach seinem Geschmack selbst zusammen), Lernen/Arbeiten mit Gleichgesinnten (peers), sozialer Kontext: Lernen im Kontakt von „Mensch zu Mensch“, „Face-to-Face“ (Anderson et al. 2012 ; Calkins/Vogt 2012).

Diese Tendenzen des neuen Lernens werden sich letztlich auch im Informations- und Rechercheverhalten der Studierenden spiegeln. Schon jetzt ist gemäss Studien erkennbar, dass die Netzgeneration nicht dasselbe Anspruchsniveau zeigt bezüglich Qualität und Quantität der Information (Porter 2011). Ferebee/Davis (2011) weisen nach, dass Informationsaufnahme- und -verarbeitung bei dieser Nutzergruppe anders abläuft als bei älteren Personen.

---

<sup>12</sup> S. Davidson/Goldberg 2012, S. 251.

<sup>13</sup> Udacity: <http://www.udacity.com/> (Stand 1. September 2012).

<sup>14</sup> Coursera: <https://www.coursera.org/> (Stand 1. September 2012).

<sup>15</sup> S. Website Coursera: <https://www.coursera.org/> (Stand 1. September 2012).

<sup>16</sup> Der Einfachheit halber wird durchgehend die männliche Form benutzt.

<sup>17</sup> So erwähnt Richtel (2012) eine Englischprofessorin an der Duke University, die ihre Studenten Blogbeiträge zur gerade aktuellen Lektüre verfassen lässt, statt einer Seminararbeit zu schreiben.

Gleichzeitig besteht doch auch ein Bedürfnis nach Face-to-Face-Beratung, was Carpenter et al. (2012) in einer Langzeituntersuchung über drei Jahre deutlich darlegen. Wie zeigt sich das Bild der Internetgeneration nun wirklich, ist sie tatsächlich so stark im virtuellen Raum präsent, nur auf die elektronischen Gadgets ausgerichtet, oder kann sie sich zwischendurch auch anderen Dingen erfreuen?

### **1. Methode**

Um dies herauszufinden, wird im Juni 2012 eine nicht repräsentative Umfrage zum Thema „Umgang mit sozialen Medien“ durchgeführt, Zielgruppe sind die sogenannten „Digital Natives“. Nach dem Zufallsprinzip werden 10 Personen ausgewählt, an einer Universitätsbibliothek und in einem Kaffeehaus (Starbucks). Dabei interessiert nicht die Art und Weise der Benutzung der sozialen Medien, sondern viel mehr die Aktivitäten bei der Nichtbenutzung. Dies führt zu folgenden drei Fragen:

- Alter, Geschlecht
- Was bedeutet für Sie „Musse“?
- Wie verbringen Sie Ihre Mussestunden?

### **Resultat**

An der Umfrage nehmen 10 Leute teil, davon sind vier Männer, fünf Frauen, eine Person macht keine Angaben zum Geschlecht. Die Altersspanne bewegt sich zwischen 16 und 29 Jahren (16jährig: 2x, 20jährig: 2x, 22jährig: 1x, 23jährig: 2x, 25jährig: 1x, 29jährig: 1x, einmal keine Angaben), neun der Befragten sprechen Deutsch, eine Person redet Englisch. Die Antworten lassen auf eine ziemlich homogene Gruppe schließen. „Musse“ bedeutet für die meisten Erholung, Abschalten, Geniessen, keine Verpflichtungen haben, Freizeit. Dementsprechend wird die Frage „Wie verbringen Sie Ihre Mussestunden?“ beantwortet; die häufigsten Nennungen sind hier: Sport/Bewegung, Freunde sehen/treffen, Garten/Natur.

Nur zwei Personen (1 Mann, zweite Person ohne Geschlechtsangabe) erwähnen eine Beschäftigung, die technischer Natur ist (Computer bzw. Surfen/PC/Facebook), wobei die Facebook-Tätigkeit sofort wieder entkräftet wird durch die Aussage, dass dies eigentlich auf lange Zeit gar nicht mehr interessant sei (hier könnte die hypothetische

Annahme gemacht werden, dass die betreffende Person ihre Mussezeit in Zukunft wohl mit etwas anderem verbringen wird).

Eine Person fasst „Musse“ in konträrem Sinne auf, als „Müssen“, „sich zu einer bestimmten Sache durchringen zu müssen“ (die Auswertung der Fragebogen findet sich im Anhang).

Diese Erkenntnisse erstaunen zumindest teilweise. Entgegen der Annahme, dass bei dieser Altersgruppe auch die Freizeitgestaltung von den (sozialen) Medien dominiert wird, erweist sich dies bei den Interviewten als falsch. Obwohl nicht repräsentativ, verdeutlicht diese Umfrage doch, dass die Jungen sehr wohl die Fähigkeiten besitzen, sich zumindest in ihrer Freizeit bewusst auszuklinken und abzuschalten.

#### *Angebote und Dienstleistungen für die Netzgeneration*

Wie lassen sich Angebote und Dienstleistungen bilden, die dem Verhalten der jungen Generation gerecht werden? Das Ziel muss es sein, physischen und virtuellen Raum zu schaffen, der das facettenreiche Spektrum von Arbeit/Studium und Freizeit/Auszeit vereint und nahtlos ineinander verzahnt. Im Folgenden werden einige Ideen präsentiert.

Der Raum soll die beiden Elemente der kreativen Schaffensperiode (Arbeit/Studium) und der Entspannung (Freizeit/Auszeit) aufnehmen, um so diese Gegenpole zu versinnbildlichen. Er soll weiter inspirations- und motivationsfördernd wirken, zum Denken einladen, dazu verleiten, Visionen zu entwickeln, für sich persönlich, für die Welt... Um diese Werte zu fördern, kann z.B. jeden Monat eine „Frage des Monats“ im Raum ausgehängt werden; geeignet sind Fragen philosophischer Art, auf die es keine eindeutige Antworten gibt.<sup>18</sup> Der Name des Raums ist Programm und wird gezielt ausgewählt, um auch hier wieder auf das Spannungsfeld Arbeit/Auszeit hinzuweisen:

- Lern-Raum
- Lehr-Raum
- Leer-Raum

---

<sup>18</sup> Die „Frage des Monats“ ist in Anlehnung an die „Annual Question“ der Denkergruppe „Edge.org“ (<http://www.edge.org>; <http://edge.org/annual-question>) (Stand 1. September 2012) entstanden. Die Inputs dazu liefern die Bibliotheksmitarbeiter oder auch die Studierenden selbst. Fragen könnten z.B. sein: Wie (wo?) findet man die Philosophie des Lebens?, Woraus besteht ein Text?

Der Lern-Raum steht dabei für das Lernen selbst, der Lehr-Raum soll die Lehre (und Forschung) symbolisieren, der Leer-Raum verdeutlicht den Prozess des „Sich-Leerens“, „Sich-Freimachens von Unnötigem“, um wieder offen zu sein für neue Inhalte.<sup>19</sup> Die Namensgebung wird in der Einrichtung des ganzen Bereichs weitergetragen.

Die in den 1990er Jahren entstehenden neuartigen Konzepte der sogenannten „Information Commons“ (IC) bzw. „Learning Commons“ (LC)<sup>20</sup> rücken die Bedürfnisse und Forderungen der Benutzer in ein neues Licht, eine neue Dimension von Informationskompetenz entwickelt sich, Kern der „Commons“ ist die Verzahnung von technischer Infrastruktur mit bibliotheksspezifischen Angeboten. Gläser (2008) beschreibt „Commons“ als eine Art gemeinsam genutzten Gemeindeacker, vergleichbar einer „Allmende“. Diese Punkte in Betracht ziehend, soll der neue Raum sowohl als Arbeits-, Computer- und Ruhezone dienen; in der Arbeitszone findet man die übliche technische Grundausstattung (inklusive WLAN), die Computerzone bietet einen umfangreichen Softwarepool zur freien Benutzung, in der Ruhezone gibt es neben Sitz- auch Liegemöglichkeiten.

Nebst der Einrichtung des Raumes stehen der persönliche Kontakt zum Gegenüber, die Vernetzung mit dem anderen, die persönliche Betreuung der Studierenden im Mittelpunkt.

Die Beratung ist auf jede einzelne Person zugeschnitten, individuell auf diese abgestimmt; dem Studenten steht auf Anfrage eine Fachperson als persönlicher Informationsberater und -coach zur Seite. Die Beratung erstreckt sich neben der allgemeinen Vermittlung von Medien- und Informationskompetenz insbesondere auf die Unterstützung im Umgang mit sozialen Medien, wobei es hier jedoch nicht um das technische Knowhow geht, sondern vielmehr die persönlichen Fähigkeiten im Vordergrund stehen, die Herausforderungen und Auswirkungen dieser Medien zu erkennen (Aus-, Abschalten, Multitasking versus konzentriertes Arbeiten...).

---

<sup>19</sup> Der Einfachheit halber wird dieser Lern-/Lehr-/Leer-Raum im Folgenden als „Raum“ bezeichnet. Die Beschriftung dieses Raums wird so gelöst, dass je eine Raumseite mit einer Bezeichnung versehen wird; dabei bleibt noch eine Wand frei (leer), frei zur Inspiration...

<sup>20</sup> Gemäss der Online-Enzyklopädie ODLIS (Reitz o. J.) wird „Information Commons“ wie folgt definiert: „A new type of technology-enhanced collaborative facility on college and university campuses that integrates library and computer application services (information, technology, and learning) in a single floor plan, often equipped with a wireless network and, in some cases, equipment for multimedia production...“

Ein weiterer wesentlicher Punkt ist die Schreibberatung (Begleitung während des Schreibprozesses bis zum Publikationsmanagement).

Der Informationscoach bietet seine Dienste zu den Öffnungszeiten der Bibliothek an, in der virtuellen Weiterführung kann dies auch als E-Mail- oder Chat-Auskunft sein.<sup>21</sup> Um die Betreuung in der virtuellen Welt möglichst nahtlos zu gewährleisten, sind verschiedene Szenarien denkbar: eine „Kommunikationsapp“ für die direkte Verbindung zum persönlichen Berater (Anforderung „24/7/365/360“),<sup>22</sup> eine App „Kurz und knapp“, welche die wichtigsten Fragen zum Recherchieren bündelt, Best Practice-Seiten zum Verfassen schriftlicher Arbeiten...

### **Ausblick**

Unser Leben gleicht einer Pendelbewegung zwischen physischer, virtueller und mobiler Welt, wird dabei zunehmend zu einem Balanceakt zwischen Rastlosigkeit und Ruhe, Ausgewogenheit und Überforderung. Das persönliche Netzwerk wächst in rasendem Tempo, begünstigt durch die sozialen Medien. Wichtiges und Unwichtiges lässt sich dabei kaum trennen; den richtigen Mix zu finden, kann sich als schwierig erweisen. Hier können sich Bibliotheken und Informationseinrichtungen neu positionieren, indem sie als Fachzentrum in allen Belangen zu sozialen Medien auftreten.

Die Metapher von der Allmende, Wiese, lässt sich noch um eine dritte Komponente erweitern, die man als „Social commons“ bezeichnen könnte: Damit wird signalisiert, dass die Informationseinrichtung grossen Wert legt auf soziale Aspekte; das Ver-Netzen, das Neben- und Miteinander der einzelnen Anspruchsgruppen wird gemeinsam und aktiv gelebt (Kontaktebene Studierende/Forscher, Peers, sowie Ebene persönlicher Berater/Studierende/Forscher). Der Begriff „Informationskompetenz“<sup>23</sup> sollte neu gedacht werden, erweitert werden um psychologische, soziale Momente. Solche Überlegungen fliessen bereits ein im „Researcher Development Framework (RDF)“, welches im Jahre 2010 als Kompetenz-Standard für Forscher entwickelt wird (Vitae

---

<sup>21</sup> Die Chat-Auskunft wird ebenfalls während der Bibliotheksöffnungszeiten betreut, E-Mail-Anfragen werden innert eines Arbeitstages erledigt. Dies bedingt natürlich eine fein organisierte Planung der personellen Ressourcen.

<sup>22</sup> Es wäre z.B. auch denkbar, eine Anwendung zu entwickeln, die sich in Standardlehrbücher integrieren lässt, damit bei aufkommenden Fragen direkt die Fachperson konsultiert werden kann (inklusive mobiler Anwendung).

<sup>23</sup> Es existieren verschiedene Standards zu Informationskompetenz; eine gute Übersicht dazu findet sich in Franke (2012).



2011 ; Research Information Network 2011). Im Wortlaut heisst es dort u.a.: „develop my/their research profile and professional networks using social media (eg blogs and LinkedIn)“.<sup>24</sup> Analog dazu ist eine Übernahme eines solchen Passus ins Deutsche denkbar. Ein weiterer nicht zu unterschätzender Punkt ist die Kommunikation und Zusammenarbeit mit der Universität und den einzelnen Fakultäten, um auf die Curricula ausgerichtete, individualisierte Inhalte und Beratungsangebote zu erstellen.

In der hochtechnisierten Zeit gewinnen die Werte zwischenmenschlicher Kommunikation im realen Leben wieder an Bedeutung. Bibliotheken können zu dieser Entwicklung beitragen, indem sie Raum bieten fürs Studium, Raum als Treffpunkt, als Ort... Loten Sie das Spannungsfeld zwischen realer und virtueller Welt aus, fördern Sie dabei auch die nötige Neugierde und Lust, die es braucht, um Fragen zu stellen, um dieselben Fragen immer wieder neu zu stellen und bis in die Tiefe zu ergründen.

---

<sup>24</sup> S. Research Information Network 2011, S. 15.

## Auswertung der Fragebogen

Person	Alter/ Geschlecht	Was bedeutet für Sie „Musse“?	Wie verbringen Sie Ihre Mussestunden?
<b>A</b>	25, männlich	Musse bedeutet für mich, sich zu einer bestimmten Sache durchringen zu müssen.	Hausarbeit, Organisatorisches
<b>B</b>	23, weiblich	Erholung, Hingabe, emotionale Zugewandtheit	Musizieren, in der Natur sein
<b>C</b>	20, männlich	mal abschalten und den Kopf frei bekommen, die Natur und das Leben genießen	Ich treibe Sport, ausserdem versuche ich, meine Freunde möglichst oft zu sehen, nach Möglichkeit verbindet man beides, indem man zusammen Sport macht. Ansonsten kochen wir gerne gemeinsam.
<b>D</b>	20, männlich	Eine Zeit ohne jegliche Verpflichtungen und frei von irgendwelchen Gedanken an Dinge, die in naher Zukunft organisiert werden müssen	Sport, Ausgang, Computer
<b>E</b>	hier leider keine Angaben	Freizeit, Erholung, die Zeit, in der man selbst bestimmt, was man macht	Filme schauen, Gitarre spielen, Singen, Tanzen, Surfen (PC, Facebook), Freunde treffen, Lesen Facebook brauche ich nicht mehr so oft. Ich pflege auch mein Profil nicht so. Am Anfang war es spannend, aber mit der Zeit interessiert einem nicht mehr, was dieser und jener den ganzen Tag so treibt. Facebook lenkt ab, vor allem auch, weil es dort eine Chatfunktion hat. So ist man immer und überall erreichbar...
<b>F</b>	23, männlich	Eine Form der Erholung, die nicht zweckgebunden ist. Also nicht: „Ich will mich entspannen, damit ich nachher dies und jenes machen mag.“ Sondern ein Geniessen um des Geniessens willen.	im Garten oder auf dem Balkon, mit oder ohne Musik, in der Sonne etwas umpflanzend
<b>G</b>	29, weiblich	Ruhe, Nicht-Müssen, Abschalten, Vergessen, in mich Ein- und Abtauchen, meine Umwelt in mich aufsaugen	bevorzugter Weise draussen in der Natur, mit viel Bewegung, Sport oder auch an einem ruhigen Ort, sitzend, lesend, schreibend, träumend, in Gedanken versunken
<b>H</b>	16, weiblich	Ich gehe in die Pfadi oder mache etwas mit Freunden, oft lese ich auch ein Buch oder schaue Fernseh, Telefonieren mache ich auch gerne	pure Entspannung
<b>I</b>	16, weiblich	Ich gehe ins Training oder Laufen; je nach Wetter bleibe ich auch zuhause	Abwechslung, Kopf von Gedanken frei machen, Spass, Freiheit
<b>J</b>	22, weiblich	Swim, go to movies, study for school	hier keine Angaben

## Literaturverzeichnis

[Anderson et al. 2012]

Anderson, Janna Quitney / Boyles, Jan Lauren / Rainie, Lee: The future impact of the Internet on higher education. Washington : Pew Research Center, 2012. URL: [http://pewinternet.org/~media/Files/Reports/2012/PIP\\_Future\\_of\\_Higher\\_Ed.pdf](http://pewinternet.org/~media/Files/Reports/2012/PIP_Future_of_Higher_Ed.pdf) (Stand 1. September 2012).

[Beloit College 2012]

Beloit College: The mindset list. 2016 list. O.O. 2012. URL: <http://www.beloit.edu/mindset/2016/> (Stand 1. September 2012).

[Brokaw 2012]

Brokaw, Leslie: To Be More Productive, Limit Interruptions. In: MITSloan Management Review, 5/11/12. o. S. URL: <http://sloanreview.mit.edu/improvisations/2012/05/11/to-be-more-productive-limit-interruptions/> (Stand 1. September 2012).

[Calkins/Vogt 2012]

Calkins, Andrew / Vogt, Kristen: Next Generation Learning. The Pathway to Possibility. O.O. 2012. URL: <http://net.educause.edu/ir/library/pdf/NGW1201.pdf> (Stand 1. September 2012).

[Carpenter et al. 2012]

Carpenter, Julie / Wetheridge, Louise / Tanner, Sophie: Researchers of Tomorrow. The research behaviour of Generation Y doctoral students. O.O. 2012. URL: <http://www.jisc.ac.uk/media/documents/publications/reports/2012/researchers-of-tomorrow.pdf> (Stand 1. September 2012).

[Davidson/Goldberg 2012]

Davidson, Cathy N. / Goldberg, David Theo: Our Digital Age. Implications for learning and its (online) institutions, E-Learning and Digital Media. Vol. 9 (2012), Nr. 3, S. 249-266. DOI: <http://dx.doi.org/10.2304/elea.2012.9.3.249> (Stand 1. September 2012).

[Ferebee/Davis 2011]

Ferebee, Susan / Davis, James Wayne: Persuading library use in technologically structured individuals. In: Library Hi Tech, Vol. 29 (2011), Nr. 4, S. 586 – 604.

[Franke 2012]

Franke, Fabian: Standards der Informationskompetenz für Studierende. In: Sühl-Strohmenger, Wilfried [Hrsg.]: Handbuch Informationskompetenz. Berlin : Walter de Gruyter [u.a.], 2012. S. 235-249.

[Gläser 2008]

Gläser, Christine: Die Bibliothek als Lernort. Neue Servicekonzepte. In: Bibliothek. Forschung und Praxis, Jg. 32 (2008), Nr. 2, S. 171-182.

[Kühne 2011]

Kühne, Martina: Servicekultur im Netzzeitalter. Zwischen Algorithmen und Intuition – Wie digitale Dienste zu sinnlichen Erlebnissen werden. Zürich: GDI, 2011.

[Porter 2011]

Porter, Brandi: Millennial Undergraduate Research Strategies in Web and Library Information Retrieval Systems. In: Journal of Web Librarianship, Vol. 5 (2011), Nr. 4, S. 267-285. DOI: <http://dx.doi.org/10.1080/19322909.2011.623538> (Stand 1. September 2012).

[Prensky 2001]

Prensky, Mark: Digital Natives, Digital Immigrants. Part II: Do they really *think* differently?. In: On the Horizon, Vol. 9 (2001), Nr. 6, S. 1-6. DOI: <http://dx.doi.org/10.1108/10748120110424843> (Stand 1. September 2012).

[Reitz o.J.]

Reitz, Joan M.: ODLIS. Online Dictionary for Library and Information Science. O.O. o.J. URL: <http://lu.com/odlis> (Stand 1. September 2012).

[Research Information Network 2011]

Research Information Network [Hrsg.]: The role of research supervisors in information literacy. A Research Information Network report. O.O. 2011. URL: [http://rinarchive.jisc-collections.ac.uk/system/files/attachments/Research\\_supervisors\\_report\\_for\\_screen.pdf](http://rinarchive.jisc-collections.ac.uk/system/files/attachments/Research_supervisors_report_for_screen.pdf) (Stand 1. September 2012).

[Richtel 2012]

Richtel, Matt: Blogs vs. Term Papers. O.O. 2012. URL: <http://www.nytimes.com/2012/01/22/education/edlife/muscling-in-on-the-term-paper-tradition.html> (Stand 1. September 2012).

[Rideout et al. 2010]

Rideout, Victoria J. / Foehr, Ulla G. / Roberts, Donald F.: Generation M<sup>2</sup>. Media in the Lives of 8- to 18-Year-Olds. Menlo Park: Henry J. Kaiser Family Foundation 2010. URL: <http://www.kff.org/entmedia/upload/8010.pdf> (Stand 1. September 2012).

[Small/Vorgan 2009]

Small, Gary / Vorgan, Gigi: iBrain. Wie die neue Medienwelt das Gehirn und die Seele unserer Kinder verändert. Stuttgart : Kreuz, 2009.

[Stone 2005]

Stone, Linda: Linda Stone's blog. O.O. 2005. URL: <http://lindastone.net/qa/continuous-partial-attention/> (Stand 1. September 2012).

[Vitae 2011]

Vitae [Hrsg.]: Researcher Development Framework (RDF). Version 2. O.O. 2011. URL: <http://www.vitae.ac.uk/CMS/files/upload/Vitae-Researcher-Development-Framework.pdf> (Stand 1. September 2012).

[White/Le Cornu 2011]

White, David S. / Le Cornu, Alison: Visitors and residents. A new typology for online engagement. In: First Monday, Vol. 16 (2011), Nr. 9. o.S. URL: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3171/3049> (Stand 1. September 2012).

[White et al. 2012]

White, David / Connaway, Lynn Silipigni / Lanclos, Donna / Le Cornu, Alison / Hood, Erin: Digital Visitors and Residents. Progress Report. O.O. 2012. URL: <http://www.jisc.ac.uk/media/documents/projects/visitorsandresidentsinterim%20report.pdf> (Stand 1. September 2012).

[Wissen o.J.]

Wissen.de. Gütersloh [u.a.] : wissenmedia in der inmediaONE] GmbH, o.J. URL: <http://www.wissen.de/> (Stand 1. September 2012).

---

# **Liste der Autoren**



**Autorenliste**

<b>Dr. Jörg Berkemeyer</b> .....	219
IHS Global Insight Ltd Bleichstraße 1 60322 Frankfurt am Mai joerg.berkemeyer@ihs.com www.ihs.com	
<b>Daniel Beucke</b> .....	207
Gruppe Elektronisches Publizieren (EPU) Georg-August-Universität Göttingen Niedersächsische Staats- und Universitätsbibliothek Göttingen 37073 Göttingen beucke@sub.uni-goettingen.de	
<b>Sylvia Brink</b> .....	307
Georg-Eckert-Institut für internationale Schulbuchforschung Celler Straße 3 38114 Braunschweig brink@gei.de	
<b>Sammy David</b> .....	149
Computer- und Medienservice / Institut für Bibliotheks- und Informationswissenschaft Humboldt-Universität zu Berlin Unter den Linden 6 10099 Berlin sammy.david@cms.hu-berlin.de	
<b>Bernd Decker</b> .....	323
RWTH Aachen University Rechen- und Kommunikationszentrum Seffenter Weg 23 52074 Aachen	
<b>Dr. Michael Eble</b> .....	135
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS Schloss Birlinghoven 53754 Sankt Augustin michael.eble@iais.fraunhofer.de	
<b>Kai Eckert</b> .....	107
Universitätsbibliothek Mannheim Schloss Schneckenhof West 68131 Mannheim	



<b>Frank Förster</b> .....	195
Christian-Albrechts-Universität zu Kiel	
Institut für Klassische Altertumskunde	
Leibnizstr. 8	
24118 Kiel	
ffoerster@gshdl.uni-kiel.de	
<b>Mark Frenzel</b> .....	75
Helmholtz-Zentrum für Umweltforschung UFZ	
Departments Umweltinformatik/Department Biozönoseforschung	
Permoserstraße 15	
04318 Leipzig	
<b>Andreas L. Fuchs</b> .....	307
Georg-Eckert-Institut für internationale Schulbuchforschung	
Celler Straße 3	
38114 Braunschweig	
afuchs@gei.de	
<b>Prof. Dr. Simone Fühles-Ubach</b> .....	337
Fakultät für Informations- und Kommunikationswissenschaften	
Fachhochschule Köln	
Gustav-Heinemann-Ufer 54	
50968 Köln	
simone.fuehles-ubach@fh-koeln.de	
<b>Rainer Gasche</b> .....	75
Karlsruher Institut für Technologie (KIT)	
Institut für Meteorologie und Klimaforschung (IMK-IFU)	
Kreuzeckbahnstrasse 19	
82467 Garmisch-Partenkirchen	
<b>Matthias Glasow</b> .....	259
Universität Rostock	
Forschungsstelle Universitätsgeschichte	
Schwaansche Straße 4	
18055 Rostock	
matthias.glasow@uni-rostock.de	
<b>Dr. Hans-Jürgen Goebelbecker</b> .....	61
Karlsruher Institut für Technologie (KIT), Bibliothek	
Kaiserstraße 12	
76131 Karlsruhe	
goebelbecker@kit.edu	

---

<b>Stefanie Grunow</b> .....	45
Legal Affairs & Licensing	
ZBW – German National Library of Economics	
Leibniz Information Centre for Economics	
Neuer Jungfernstieg 21	
20354 Hamburg	
s.grunow@zbw.eu	
 <b>Manfred Hauer</b> .....	 235
AGI - Information Management Consultants	
Mandelring 238 b	
67433 Neustadt / Weinstrasse	
Manfred.Hauer@agi-imc.de	
 <b>Brigitte Hausstein</b> .....	 45
Projektmanagerin da ra	
Registration Agency für Social and Economic Data	
GESIS – Leibniz-Institute for Social Sciences (GESIS)	
Schiffbauerdamm 19	
10117 Berlin	
Brigitte.Hausstein@gesis.org	
 <b>Lambert Heller</b> .....	 295
Fachreferent Wirtschaftswissenschaften und Wirtschaftsinformatik	
TIB/UB Hannover	
Königsworther Platz 1B	
30167 Hannover	
<a href="http://wikify.org/">http://wikify.org/</a>	
 <b>Roderich Henry</b> .....	 307
Georg-Eckert-Institut für internationale Schulbuchforschung	
Celler Straße 3	
38114 Braunschweig	
henry@gei.de	
 <b>Prof. Dr. Hans-Christoph Hobohm</b> .....	 93
FH-Potsdam, FB Informationswissenschaften	
Friedrich-Ebert-Str. 4	
14467 Potsdam	
hobohm@fh-potsdam.de	

<b>Julia Iwanowa</b> .....	149
Computer- und Medienservice / Institut für Bibliotheks- und Informationswissenschaft Humboldt-Universität zu Berlin Unter den Linden 6 10099 Berlin julia.iwanowa@ibi.hu-berlin.de	
<b>Maxi Kindling</b> .....	149
Computer- und Medienservice / Institut für Bibliotheks- und Informationswissenschaft Humboldt-Universität zu Berlin Unter den Linden 6 10099 Berlin maxi.kindling@hu-berlin.de	
<b>Sebastian Kirch</b> .....	135
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS Schloss Birlinghoven 53754 Sankt Augustin	
<b>Jens Klump</b> .....	75
Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ Zentrum für Geoinformationstechnologie Telegrafenberg, 14473 Potsdam	
<b>Olaf Kolditz</b> .....	75
Helmholtz-Zentrum für Umweltforschung UFZ Departments Umweltinformatik/Department Biozönoseforschung Permoserstraße 15 04318 Leipzig	
<b>Dr. Ralf Kunkel</b> .....	75
Forschungszentrum Jülich GmbH Institut für Bio- und Geowissenschaften – Agrosphäre (IBG-3) 52425 Jülich r.kunkel@fz-juelich.de	
<b>Karsten Labahn</b> .....	259
Universitätsbibliothek Rostock Albert-Einstein-Straße 6 18059 Rostock karsten.labahn@uni-rostock.de	

<b>Prof. Dr. Viktor Mayer-Schönberger</b> .....	9
Oxford Internet Institute University of Oxford 1 St Giles Oxford OX1 3JS United Kingdom viktor.ms@oii.ox.ac.uk	
<b>Mirosław Mainusch</b> .....	323
RWTH Aachen University Hochschulbibliothek, IT-Dezernat Templergraben 61 52056 Aachen mainusch@bth.rwth-aachen.de	
<b>Lars Müller</b> .....	93
FH-Potsdam, FB Informationswissenschaften Friedrich-Ebert-Str. 4 14467 Potsdam lars.mueller@fh-potsdam.de	
<b>Frank Neidl</b> .....	75
Karlsruher Institut für Technologie (KIT) Institut für Meteorologie und Klimaforschung (IMK-IFU) Kreuzheckbahnstrasse 19 82467 Garmisch-Partenkirchen	
<b>Michael Nolde</b> .....	195
Christian-Albrechts-Universität zu Kiel Geographisches Institut (Sektion Geographie) Ludewig-Meyn-Str. 14 24118 Kiel	
<b>Dr. Wiebke Oeltjen</b> .....	225
Universität Hamburg Schlüterstr. 70 20146 Hamburg wiebke.oeltjen@uni-hamburg.de	
<b>Felix Ostrowski</b> .....	271
literarymachine.net	

<b>Heinz Pampel</b> .....	61
Deutsches GeoForschungsZentrum GFZ	
Bibliotheks- und Informationsdienste (LIS)	
Telegrafenberg	
14473 Potsdam	
pampel@gfz-potsdam.de	
<b>PD Dr. med. Alberto Perez Bouza</b> .....	323
Universitätsklinikum Bonn, Institut für Pathologie	
Sigmund-Freud-Str 25	
53127 Bonn	
<b>Adrian Pohl</b> .....	271
Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen, hbz	
Jülicher Str. 6	
50674 Köln	
pohl@hbz-nrw.de	
<b>Dominique Ritze</b> .....	107
Universitätsbibliothek Mannheim	
Schloss Schneckenhof West	
68131 Mannheim	
dominique.ritze@bib.uni-mannheim.de	
<b>Dr. Dominik Schmitz</b> .....	323
RWTH Aachen University	
Hochschulbibliothek, IT-Dezernat	
Templergraben 61	
52056 Aachen	
<b>Dr. Silke Schomburg</b> .....	17
Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz)	
Jülicher Straße 6	
50674 Köln	
schomburg@hbz-nrw.de	
<b>Andrea Siegberg</b> .....	247
Abteilung Informationsmanagement	
Fraunhofer Zentrale	
Schloss Birlinghoven	
53754 Sankt Augustin	
andrea.siegberg@zv.fraunhofer.de	

<b>Olaf Siegert</b> .....	29
ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften	
Leibniz Informationszentrum Wirtschaft	
Neuer Jungfernstieg 21	
20354 Hamburg	
O.Siegert@zbw.eu	
<b>Doris Škarić</b> .....	165
Bayerische Staatsbibliothek	
80328 München	
doris.skarić@bsb-muenchen.de	
<b>Jürgen Sorg</b> .....	75
Forschungszentrum Jülich GmbH	
Institut für Bio- und Geowissenschaften – Agrosphäre (IBG-3)	
52425 Jülich	
<b>Stefanie Sorge</b> .....	119
Hochschule Biberach, Bibliothek	
Karlstr. 11	
88400 Biberach	
<b>Robert Stephan</b> .....	259
Universitätsbibliothek Rostock	
Albert-Einstein-Straße 6	
18059 Rostock	
robert.stephan@uni-rostock.de	
<b>Robert Strötgen</b> .....	307
Georg-Eckert-Institut für internationale Schulbuchforschung	
Celler Straße 3	
38114 Braunschweig	
stroetgen@gei.de	
<b>Prof. Dr. Bernhard Thalheim</b> .....	195
Christian-Albrechts-Universität zu Kiel	
Institut für Technologie der Informationssysteme	
Hermann-Rodewald-Str. 3	
24118 Kiel	
<b>Paul Vierkant</b> .....	61
Humboldt-Universität zu Berlin	
Institut für Bibliotheks- und Informationswissenschaft (IBI)	
Unter den Linden 6	
10099 Berlin	
paul.vierkant@hu-berlin.de	

<b>Sven Vlaeminck</b> .....	29
ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften	
Leibniz Informationszentrum Wirtschaft	
Neuer Jungfernstieg 21	
20354 Hamburg	
s.vlaeminck@zbw.eu	
<b>Michaela Voigt</b> .....	149
Computer- und Medienservice /	
Institut für Bibliotheks- und Informationswissenschaft	
Humboldt-Universität zu Berlin	
Unter den Linden 6	
10099 Berlin	
voigtmic@cms.hu-berlin.de	
<b>Georg von Nessler</b> .....	219
IP-ShareMedia	
Neuer Wall 4	
60594 Frankfurt am Main	
nessler@ip-sharemedia.de	
www.IP-ShareMedia.de	
<b>Dr. Jürgen Warmbrunn</b> .....	179
Forschungsbibliothek	
Herder-Institut für historische Ostmitteleuropaforschung -	
Institut der Leibniz-Gemeinschaft	
Gisonenweg 5-7	
35037 Marburg	
juergen.warmbrunn@herder-institut.de	
<b>Anne-Katharina Weilenmann</b> .....	351
MSc Bibliotheks- und Informationsmanagement	
Stettenerstrasse 165	
CH-8207 Schaffhausen	
BIBLINK.ch	
weilenmann@onlinetravel.ch	
http://www.biblink.ch	
<b>Dr. med. Saskia Eva Elisabeth Westphal</b> .....	323
RWTH Aachen University, Institut für Pathologie	
Pauwelsstr. 30	
52074 Aachen	

<b>Thomas Wetzel</b> .....	93
FH-Potsdam, FB Informationswissenschaften	
Friedrich-Ebert-Str. 4	
4467 Potsdam	
wetzel@fh-potsdam.de	



---

# **Sponsoren und Aussteller**



Wir danken folgenden **Sponsoren** für Ihre freundliche Unterstützung:

**Hauptsponsor:** EBSCO Information Services GmbH

**Weitere Sponsoren:**

Duncker & Humblot GmbH  
Elsevier B.V.  
GALE Cengage Learning  
GLOMAS Deutschland GmbH  
MIRA Consulting GmbH  
Georg Olms Verlagsbuchhandlung  
Springer-Verlag GmbH  
Swets Information Services GmbH

**Aussteller:**

CAS American Chemical Society	Neschen AG
ciando GmbH	Walter Nagel GmbH & Co. KG
De Gruyter GmbH & Co. KG	Ovid Technologies GmbH
Dietmar Dreier Wissenschaftliche Verlagsbuchhandlung GmbH	SAM IEEE Standards and More GmbH & Co. KG
Elsevier B.V.	Schweitzer Fachinformationen
HARRASSOWITZ GmbH & Co. KG	Springer - Verlag GmbH
IHS Global Insight Ltd	Swets Information Services GmbH
Image Access	Thomson Reuters
Lehmanns Media GmbH	Verlagshaus Monsenstein und Vannerdat OHG
MIRA Consulting GmbH	VUB Printmedia GmbH
Missing Link Versandbuchhandlung eG	



1. **Naturwissenschaft und Technik – nur für Männer?  
Frauen mischen mit!**  
Auswahl-Bibliographie Wissenschaftlerinnen (1999), 28 Seiten  
ISBN: 978-3-89336-246-2
4. **Schweißen & Schneiden**  
Wissenschaftliche Veröffentlichungen des Forschungszentrums Jülich  
(1997), 16 Seiten  
ISBN: 978-3-89336-208-0
5. **Verzeichnis der wissenschaftlich-technischen Publikationen**  
des Forschungszentrums Jülich  
Januar 1993 - Juli 1997 (1997), ca. 100 Seiten  
ISBN: 978-3-89336-209-7
6. **Biotechnologie**  
Wissenschaftliche Veröffentlichungen der Institute für Biotechnologie  
des Forschungszentrums Jülich  
Januar 1992 - Juni 1997 (1997), 48 Seiten  
ISBN: 978-3-89336-210-3
7. **Verzeichnis der wissenschaftlich-technischen Publikationen**  
des Forschungszentrums Jülich  
1997 bis 1999 (2000), 52 Seiten  
ISBN: 978-3-89336-260-8
8. **Kompodium Information**  
Teil I: Archive, Bibliotheken, Informations- und Dokumentationseinrichtungen  
Teil II: Ausbildungsstätten, Fort- und Weiterbildungsaktivitäten, Informations-  
dienste, Presse- und Nachrichtenagenturen, Verlagswesen und Buchhandel,  
Einrichtungen des Patent- und Normungswesen, Publikationen  
von G. Steuer (2001), 1130 Seiten  
ISBN: 978-3-89336-286-8
9. **Die Zukunft des wissenschaftlichen Publizierens**  
Der Wissenschaftler im Dialog mit Verlag und Bibliothek  
Jülich, 28. bis 30. November 2001. 40 Jahre Zentralbibliothek. Konferenz und  
Firmenausstellung  
Tagungsprogramm und Kurzfassungen (2001), 50 Seiten  
ISBN: 978-3-89336-292-9
10. **Die Zukunft des wissenschaftlichen Publizierens**  
Der Wissenschaftler im Dialog mit Verlag und Bibliothek  
Jülich, 28. - 30.11.2001. Tagungsprogramm und Vorträge (2002), 184 Seiten  
ISBN: 978-3-89336-294-3 (broschiert)  
ISBN: 978-3-89336-295-0 (CD)

11. **Bibliometric Analysis in Science and Research**  
Applications, Benefits and Limitations  
2<sup>nd</sup> Conference of the Central Library, 5 – 7 November 2003, Jülich, Germany  
Conference Proceedings (2003), 242 pages  
ISBN: 978-3-89336-334-6
12. **Bibliometrische Analysen – Daten, Fakten und Methoden**  
Grundwissen Bibliometrie für Wissenschaftler, Wissenschaftsmanager,  
Forschungseinrichtungen und Hochschulen  
von R. Ball, D. Tunger (2005), 81 Seiten  
ISBN: 978-3-89336-383-4
13. **VIRUS – Sicher im Netz?**  
2. Internationale Konferenz zur Virtuellen Bibliothek des Goethe-Instituts  
Brüssel  
herausgegeben von R. Ball, C. Röpke, W. Vanderpijpen (2005), 137 Seiten mit  
beiliegender CD-ROM  
ISBN: 978-3-89336-377-3
14. **Knowledge eXtended**  
Die Kooperation von Wissenschaftlern, Bibliothekaren und IT-Spezialisten  
3. Konferenz der Zentralbibliothek, 2. – 4. November 2005 Jülich  
Vorträge und Poster (2005), 392 Seiten  
ISBN: 978-3-89336-409-1
15. **Qualität und Quantität wissenschaftlicher Veröffentlichungen**  
Bibliometrische Aspekte der Wissenschaftskommunikation  
von M. Jokić, R. Ball (2006), 186 Seiten  
ISBN: 978-3-89336-431-2
16. **Bibliotheken in Singapur**  
von B. Mittermaier (2006), 114 Seiten  
ISBN: 978-3-89336-449-7
17. **Libraries in Singapore**  
by B. Mittermaier (2007), 112 pages  
ISBN: 978-3-89336-450-3
18. **WissKom 2007: Wissenschaftskommunikation der Zukunft**  
4. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,  
6. - 8. November 2007, Beiträge und Poster  
hrsg. von R. Ball (2007), 300 Seiten  
ISBN: 978-3-89336-459-6
19. **Bibliometrische Verfahren und Methoden als Beitrag zu  
Trendbeobachtung und –erkennung in den Naturwissenschaften**  
von D. Tunger (2009), 311 Seiten  
ISBN: 978-3-89336-550-0

20. **WissKom 2010: eLibrary – den Wandel gestalten**  
5. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,  
8. – 10. November 2010, Proceedingsband  
hrsg. von B. Mittermaier (2010), 389 Seiten  
ISBN: 978-3-89336-668-2
21. **Vernetztes Wissen – Daten, Menschen, Systeme**  
6. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,  
5. – 7. November 2012, Proceedingsband  
hrsg. von B. Mittermaier (2012), 379 Seiten  
ISBN: 978-3-89336-821-1



## Vernetztes Wissen – Daten, Menschen, Systeme

Informations- und Wissensvermittlung verlagern sich immer stärker in die digitale Welt. Möglich wird dies nicht zuletzt durch die voranschreitende Durchdringung aller Lebensbereiche durch das Internet. Wissen wird mehr und mehr zu vernetztem Wissen.

Die Jülicher Konferenz WissKom2012 thematisiert die Anpassung an diese Entwicklung und ihre Mitgestaltung durch innovative Bibliotheksdienstleistungen.

Der Konferenztitel „Vernetztes Wissen: Daten, Menschen, Systeme“ deutet die wechselseitige Vernetzung unter- und miteinander an. Ziel ist, vorhandene Insellösungen zu verbinden und neue Konzepte für inhärent vernetzte Strukturen zu entwickeln.